JOURNAL OF APPLIED PHYSICS

JOURNAL OF APPLIED PHYSICS

Vol. 25, No. 3 MARCH, 1954

In This Issue

	Page
Period and Amplitude of the Van Der Pol Limit Cycle — Edward Fisher	273
Response of a Pulsed Geiger Tube — Herbert B. Rosenstock	275
Potential Distribution and Prevention of a Space-Charge-Induced Minimum Between a Plane Secondary Electron Emitter and Parallel Control Grid — George C. Sponsler	282
Methods of Processing Silver-Magnesium Secondary Emitters for Electron Tubes — Paul Rappaport	288
Symmetry as a Factor in Finite Difference Approximations — C. M. Fowler	293
Approximations in Linear Viscoelasticity Theory: Delta Function Approximations — Herbert Leaderman	294
RLC Canonic Forms — F. M. Reza	297
Resonant Cavity Study of Semiconductors — Hsi-Teh Hsieh, James M. Goldey, and Sanborn C. Brown	302
16 Percent Aluminum-Iron Alloy Cold Rolled in the Order-Disorder Temperature Range — Joseph F. Nachman and William J. Buehler	307
Diode Theory in the Light of Hole Injection — John A. Swanson	314
Boundary Disturbances in High-Explosive Shock Tubes — R. G. Shreffler and R. H. Christian	324
Arcing at Electrical Contacts on Closure. Part IV. Activation of Contacts by Organic Vapor — L. H. Germer	332
Use of Restricted Variational Principles for the Solution of Differential Equations — Philip Rosen	336
Apparent Density of Thin Evaporated Films — M. S. Blois, Jr. and L. M. Rieser, Jr.	338
Noise in Semiconductors: Spectrum of a Two-Parameter Random Signal — Stefan Machlup	341
Radiation Damage Effects on Order-Disorder in Nickel-Manganese Alloys — Lewis R. Aronin	344
Experimental Study of Low-Frequency Effects on the Dynamic Modulus of a Buna-N Rubber — Allen Q. Hutton and A. W. Nolle	350
Estimation of Correlation Coefficients from Scatter Diagrams — George R. Sugar	354
Fast Time Analysis of Intermittent Point-to-Plane Corona in Air. II. The Positive Pre-Onset Streamer Corona — M. R. Amin	358
Diffusion of Gold into Copper — A. B. Martin, R. D. Johnson, and Frank Asaro	364
General Theory of Plane-Wave Scattering from Finite, Conducting Obstacles with Application to the Two-Antenna Problems — J. E. Storer and J. Sevick	369
On a Nonlinear Diffusion Equation Applied to the Magnetization of Saturable Reactors — Shou-Hsien Chow	377
Cathode Effects in the Dielectric Breakdown of Liquids — J. K. Bragg, A. H. Sharbaugh, and R. W. Crowe	382
On the Electric Strengths of Aliphatic Hydrocarbons — R. W. Crowe, J. K. Bragg, and A. H. Sharbaugh	392
Normal Reflection of Shock Waves from Moving Boundaries — Ralph A. Alpher and Robert J. Rubin	395
Heat Transfer Measurements at Sodium-Stainless Steel Interface — James W. Moyer and William A. Riemen	400
Low-Temperature Internal Friction and Elasticity Effects in Vitreous Silica — M. E. Fine, H. Van Duyne, and Nancy T. Kenney	402
Letters to the Editor:	
Shot Dependence of *p-n* Junction Phototransistor Noise — A. Slocum and J. N. Shive	406
On an Indeterminate Integral in Einstein's Theory of the Viscosity of a Suspension — M. Mooney	406
Remark on Stability — Kenneth S. Miller	407
Titanium Observed by Electron Diffraction — Shigeto Yamaguchi	408
Index to Advertisers	xxiv

ELMER HUTCHISSON, *Editor*

EARLE C. GREGG, *Assistant Editor*

Board of Editors
MAURICE A. BIOT
RAY E. BOLZ
WALLACE J. ECKERT
ROBERT D. HEIDENREICH
LOUIS A. PIPES
ROMAN SMOLUCHOWSKI
W. R. SMYTHE
ARTHUR V. TOBOLSKY
ALVIN M. WEINBERG

Publication of the American Institute of Physics

OFFICERS
GEORGE R. HARRISON, *Chairman*
GEORGE B. PEGRAM, *Treasurer*

GOVERNING BOARD
J. W. BUCHTA
KARL K. DARROW
G. J. DIENES
S. A. GOUDSMIT
GEORGE R. HARRISON
FREDERICK V. HUNT
DEANE B. JUDD
HUGH S. KNOWLES
C. C. LAURITSEN
WILLIAM F. MEGGERS
PHILIP M. MORSE
BRIAN O'BRIEN
GEORGE B. PEGRAM
I. I. RABI
DUANE ROLLER
FREDERICK SEITZ
WILLIAM SHOCKLEY
RICHARD M. SUTTON
J. H. VAN VLECK
MARK W. ZEMANSKY

Consultant on Printing
MELVIN LOOS

ADMINISTRATIVE STAFF
HENRY A. BARTON, *Director*
WALLACE WATERFALL, *Executive Secretary*
THEODORE VORBURGER, *Advertising Manager*
RUTH F. BRYANS, *Publication Manager*
EDITH I. NEFTEL *CirculationManager*
KATHRYN SETZE *Assistant Treasurer*

The JOURNAL OF APPLIED PHYSICS, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. The JOURNAL OF APPLIED PHYSICS publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. Unless otherwise stated, submission of a manuscript is a representation that it has been neither copyrighted, published, nor submitted for publication elsewhere. The authors' institutions are requested to pay a **publication charge** of $8 per page which, if honored, entitles them to 100 reprints without further charge. Instructions will be sent with galley proofs.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members of the American Institute of Physics and Affiliated Societies	$10.00	$11.00
To all others	12.00	13.00

Back Number Prices

Yearly back number rate when complete year is available: $14.00. Single copies: $1.25 each.

Changes of address should be sent to the American Institute of Physics: **corrected proofs** should be sent to the Publication Manager, American Institute of Physics. **Advertising rates** supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 25, Number 3 March, 1954

The Period and Amplitude of the Van Der Pol Limit Cycle

EDWARD FISHER
General Precision Laboratory, Pleasantville, New York
(Received June 1, 1953; revised manuscript received October 19, 1953)

The amplitude and period of the limit cycle of the Van Der Pol equation, $\ddot{y}+y=\nu\dot{y}(1-y^2)$, are found for all values of ν by joining graphically the results of solutions about $\nu=0$ and $\nu=\infty$.

THE limit cycle of the Van Der Pol equation,

$$\frac{d^2y}{dt^2}+\nu(y^2-1)\frac{dy}{dt}+y=0, \tag{1}$$

can be determined for sufficiently small ν by the Lindstedt method.[1] This consists in assuming for y and the period, T, power series in ν, and determining the coefficients of the series by substituting in (1) and integrating the resulting system of differential equations. The method can be considerably simplified by noticing that it amounts to requiring y to have the form

$$y(t)=\cos t\sum_{k=0}^{\infty}a_k\mu^{2k}+\mu\sin^2 t\sum_{k=0}^{\infty}b_k\mu^{2k}, \tag{2}$$

where a_k and b_k are polynomials of degree $2k$ in $\sin^2 t$, and that the form of (1) requires T to be a power series in ν^2. Even with this simplification, however, the method is too laborious to carry far enough for our present purpose.

Therefore, write $y=Ax$, where A is the amplitude of the limit cycle, and transform to new variables $z=1-x^2$, $E=(\frac{1}{4})[\dot{x}(x)+\dot{x}(-x)]^2$, that is, $\sqrt{E}$ is the even part of $\dot{x}$ as a function of x. Then, from (1),

$$(\nu/2x)(A^2x^2-1)\sqrt{E}=\frac{d}{dz}\left[E\left(1-\frac{dz}{dE}\right)^{\frac{1}{2}}\right] \tag{3}$$

and

$$A^2=\int_0^1\sqrt{E}dx\Big/\int_0^1 x^2\sqrt{E}dx,\quad T=8\int_0^1\frac{d\sqrt{E}}{dz}dx. \tag{4}$$

Equations (3) and (4) show that E, A, and T are even functions of ν. Then we assume

$$E=f_0(z)+\nu^2f_1(z)+\nu^4f_2(z)\cdots,\quad A^2=a_0+a_1\nu^2\cdots. \tag{5}$$

To determine the form of the f's we substitute in (1) the series

$$\dot{x}(x)=\sum_{k=0}^{\infty}\nu^kg_k(x).$$

Inspection of the resulting recurrence relations shows that g_{2k} has the form, $(1-x^2)^{\frac{1}{2}}$ (polynomial of $2k$th degree in x^2). From this and from the form imposed by (3) on E in the neighborhood of $z=1$ it follows that

$$\begin{gathered}f_0=z,\quad f_1'(z)=c_1z(1-z),\\ f_2'(z)=(c_2z+c_3z^2+c_4z^3)(1-z),\quad f_1(0)=0=f_2(0),\end{gathered} \tag{6}$$

where the c's are constants.

Substituting (5) and (6) in (3) gives

$$c_1=1,\quad c_2=1/36,\quad c_3=11/36,\quad c_4=-\tfrac{1}{2}. \tag{7}$$

Then, substituting (5), (6), and (7) in (4) gives

$$\begin{aligned}T&=2\pi+\pi\nu^2/8-(5\pi/1536)\nu^4\cdots,\\ A^2&=\frac{1+\nu^2/12+0.0001049\nu^4\cdots}{\frac{1}{4}+(7\nu^2)/384+0.0002965\nu^4\cdots}.\end{aligned} \tag{8}$$

In (8), A^2 is not reduced to a power series in ν^2 because this would not increase the range of convergence—and might decrease it.

For large ν the asymptotic solution of (1) has been

[1] N. Minorsky, *Introduction to Non-linear Mechanics* (J. W. Edwards, 1947), Chap. XI.

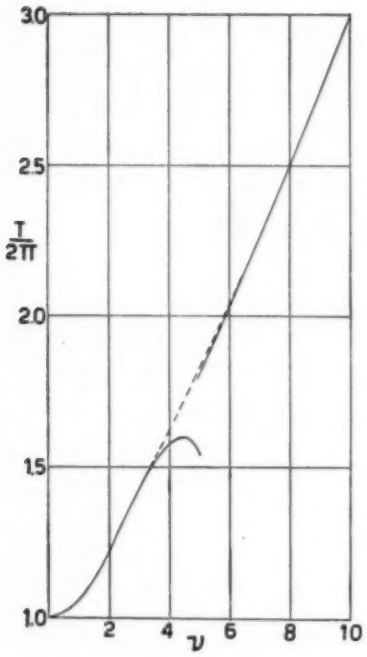


FIG. 1. The period of the limit cycle.

given by Dorodnitsin.[2] His formulas are

$$T=1.61371\nu+7.0143\nu^{-\frac{1}{3}} - (22/9)(\ln\nu)/\nu+0.0087/\nu+0(\nu^{-4/3}),$$
$$A=2+.7794\nu^{-4/3}-(16/27)(\ln\nu)/\nu^2 - 0.8762/\nu^2+0(\nu^{-8/3}), \quad (9)$$

where 0 means "terms of the order of."

Equations (8) and (9) are plotted in Figs. 1 and 2. In Fig. 1 the sharp drop in $T/2\pi$ from $\nu=4$ to 5 is considered unphysical and explained by failing convergence of the series in (8). Therefore (8) and (9) are joined smoothly by the dashed line from $\nu=3$ to 7. Similarly in Fig. 2 solutions (8) and (9) are joined graphically from $\nu=2$ to 5.

This joining of (8) and (9) is verified approximately in two ways. First, Shohat[3] has solved (1) with the Lindstedt method by a power series in the parameter $\epsilon=\nu/(1+\nu)$. His formulas are

$$A=2+\epsilon^2/96\cdots,$$
$$2\pi/T=(1/\nu)[\epsilon+\epsilon^2+(15/16)\epsilon^3+(13/16)\epsilon^4\cdots]. \quad (10)$$

[2] A. A. Dorodnitsin, Priklad. Mat. Mehan. **11**, 313 (1947).
[3] J. Shohat, J. Appl. Phys. **15**, 568 (1944).

Although he cites agreement with experiment for all ν, his derivation is unsatisfactory and, in fact, the series in (10) cannot converge everywhere because they are not even functions of ν. The question of convergence of such Lindstedt series is very difficult; although it is customary to assume convergence, it is known[4] that they can diverge. From (10) $T/2\pi=1.29$, 1.53, and 1.77 at $\nu=2$, 3, and 4, respectively.

Second, the solution of (3) can be approximated by a polynomial. For simplicity assume $(E)^{\frac{1}{2}}=z^{\frac{1}{2}}+c_1z^{\frac{3}{2}}+c_2z^{5/2}$. Then, for E to satisfy (3) in the neighborhood of $z=0$ and $z=1$ we must have $c_1=(\nu/6)^2(A^2-1)^2$,

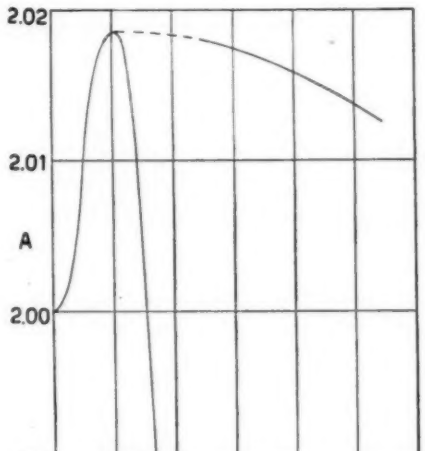


FIG. 2. The amplitude of the limit cycle.

$c_2=-\frac{1}{5}-(\nu^2/60)(A^2-1)^2$. From c_1 and c_2, with $A=2$, and Eqs. (4)

$$T/2\pi=\tfrac{5}{8}+3\nu^2/32, \quad A^2=(8/5)(28+3\nu^2)/(12+\nu^2),$$

and hence, for $\nu=2$, 3, and 4, $T/2\pi=1$, 1.47, and 2.1, and $A=2.00$, 2.04, and 2.08, respectively.

[4] H. Poincaré, *Méthodes nouvelles de la mécanique céleste*, (Gauthier-Villars et Fils, 1893), Chap. XIII.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 3 MARCH, 1954

Response of a Pulsed Geiger Tube

HERBERT B. ROSENSTOCK
Naval Research Laboratory, Washington, D. C.
(Received May 28, 1953)

For the purpose of extending the usefulness of a Geiger counter to higher radiation intensities, it has been suggested that the voltage applied to its electrodes be a periodically repeated pulse rather than the usual dc. If the duration of the pulse is of the order of the spread time of the discharge in the tube, the output current *vs* radiation intensity relation will differ from the usual linear one, largely because the resulting current surge may not reach its maximum before the end of the applied pulse. This relation will therefore depend, among other factors, on the not fully understood mechanism by which the discharge spreads. Using probability considerations, we here calculate this relation as a function of the response of the tube under ordinary operation. If the discharge spreads rapidly (e.g., exponentially) initially, an approximate general relationship not involving the detailed spread mechanism may be established. The response to be expected in two types of tubes is discussed in detail, using models for the spread mechanism which are mathematically simplified though probably physically adequate.

I. INTRODUCTION

THE Geiger tube is used as a radiation detector because an ionizing particle impinging on it will cause it to produce a surge of current. This current reaches its maximum rather quickly ("rise time" of about 1 μsec), but the tube is unable to respond to another particle for a much longer time ("dead time" of about 100 μsec).[1] For the purpose of measuring radiation intensities so large that a Geiger tube may receive more than one particle in the time interval the length of its dead time, it has been suggested[2] that the voltage applied to its electrodes be a short pulse of duration T, repeated at regular, comparatively large, intervals, N times per second, rather than the usual constant voltage. T is to be shorter than the dead time of the tube, so that the response will be at most one surge of current per applied pulse. If T is of the order of magnitude of the duration of the current surge, then the output current resulting from one applied voltage pulse will depend on the time available to the ionizing particles to multiply before the voltage on the electrodes is cut off. In addition, the current will depend on the shape of the applied pulse and on the rate of spread of the discharge in the tube at a given voltage (i.e., on the characteristics of the tube).

In this note we attempt to derive quantitative relations between these above quantities: the response of the tube to a single particle under ordinary operating conditions (i.e., constant electrode potential), the response of the tube when operated by the pulsed potential described above (the output current is usually integrated electronically, so that we may properly deal with average quantities), and radiation intensity.

[1] A general survey is given, e.g., in H. Friedman, Proc. Inst. Radio Engrs. **37**, 791 (1949).
[2] S. W. Lichtman, Nucleonics **11**, 22 (1953); B. B. Rossi and H. H. Staub, *Ionization Chambers and Counters*, (McGraw-Hill Book Company, Inc., New York, 1949) pp. 117–122.

II. GENERAL TREATMENT

Let the applied voltage pulse last for T seconds, and let $g(T-t)$ be the total charge that flows through the tube if a particle comes in a time t after the start of the pulse. $g(T-t)$ itself will depend on both the tube characteristics and the pulse shape; we are not now concerned with the exact nature of this dependence. Let us merely note that $g(T-t)$ will for physical reasons increase monotonically from some value $g(T)=g_{\max}$ at $t=0$ to $g(0)=0$ at $t=T$. Let us calculate the average output of the tube as a function of the intensity of incident radiation, assuming that the incoming particles are randomly distributed in time.

Let there be an average number m of particles incident per second, and $\alpha=mT$ particles incident per voltage pulse. We must ask ourselves two questions: First, if on the average there are α incidents per interval of fixed length, what is the probability $p(\alpha,k)$ that there occur exactly k incidents in one particular interval? Second, given k incidents per interval, what is the average total charge F_k that these k particles will cause to flow across the tube? Since k can be any number from one to infinity, the average charge per pulse will be

$$\sum_{k=1}^{\infty} p(\alpha,k)F_k$$

and the average charge per second

$$E(m)=N\sum_{k=1}^{\infty} p(\alpha,k)F_k, \tag{1}$$

N being the number of applied voltage pulses per second. The answer to the first question, the calculation of $p(\alpha,k)$, is the well-known[3] Poisson distribution

$$p(\alpha,k)=\alpha^k e^{-\alpha}/k! \tag{2}$$

so that

$$E(m)/N=\sum_{k=1} F_k\alpha^k e^{-\alpha}/k!. \tag{3}$$

[3] H. Cramer, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, 1946), Chapter 16.

The second problem, the evaluation of the average response F_k of the tube to k particles arriving at times $t_1, t_2, \cdots, t_k$ from the knowledge that a particle arriving alone at time t will cause a response $g(T-t)$, cannot be solved in general unless g is known explicitly and unless the physical processes responsible for the spread of the discharge are understood. It may be remarked that quite different processes seem to be the essential ones in tubes of different construction.[4,5]

The rest of this paper is devoted mainly to the evaluation of F_k, and from this of $E(m)$ by the use of (3), with various assumptions concerning the mechanism of spread of the discharge. If it is assumed that $g(T-t)$ is a rapidly rising function (exponentially, say) then it is found that E can, in fact, be evaluated approximately, and that a relationship between E and g, good for low radiation intensities, can be established without specifying $g(T-t)$ exactly. This is done in Sec. IIIA. Then we compute g and E for two specific assumptions about the discharge mechanism. In Sec. IIIB we consider spread by photoelectric effect on the cathode (this leads to initially rapid spread, so that Sec. IIIA might apply); and in Sec. IIIC we consider spread by photoelectric effect in the gas, which leads to an average velocity of linear spread along the wire. In both Sec. IIIB and IIIC, the weak part of the calculation is the evaluation of $g(t)$ from the physical assumptions. Once this is done, the calculation of F_k and E is a statistical and quite unambiguous one. That is why in Sec. IIIA we try to carry the work as far as possible without specific physical assumptions. If physical assumptions are made, they have to be rather simple ones to make the calculation feasible, and it is known that the processes taking place in a Geiger tube are rather complicated. It is fortunate that it is the gross shape of $g(T-t)$, rather than the detailed features which disappear in the averaging process, which is important in determining F_k and E. We, therefore, believe that the forms found for $E(m)$ are substantially correct even though we would not recommend that our expressions for $g(T-t)$ be taken too seriously.

SECTION III

A. Approximation for Rapid Multiplication

In Sec. II, we established that $g(T-t)$ is an increasing function of $T-t$. Let us now assume that it is a rapidly increasing function of $T-t$. Specifically, let us assume that a particle multiplies so fast that if another particle comes in even a small fraction of T seconds after the first, the effect of the second particle will be negligible when added to the daughter particles of the first one. That is to say, only the first of all the k particles that come in during one pulse is assumed to have an effect. With rapid spread this is likely to be nearly true as long as k is not too large. We will then have

$$F_k=\int_0^T q(l,t;k,T)g(T-t)dt, \tag{4}$$

where $q(n,t;k,T)dt$ is the probability that the nth particle comes in in dt at t, given that a total of k particles will come in $0<t<T$. To calculate $q(l,t;k,T)$, note first of all that the probability of one particle not arriving before t is $(T-t/T)$ (remember the particle is assumed to be certain to arrive before T). The probability of none of the k particles arriving before T is then $(T-t/T)^k$. This must then be multiplied by the probability $dt/(T-t)$ of one particle arriving in the interval dt thereafter. But any one of the k particles will do for this, so we multiply by k. Thus,

$$q(1,t;k,T)dt=\frac{k}{T}\left(\frac{T-t}{T}\right)^{k-1}dt. \tag{5}$$

If we substitute formula (5) into (4) and (3) and sum the series over k we get

$$E(m,T)=Nm\int_0^T g(T-t)e^{-mt}dt. \tag{6}$$

It is instructive to consider the behavior of Eq. (6) in the limiting cases of very small and very large α or m. For very small α, only the first term in the series (3) is significant and we get

$$E(m,T)=Nm\int_0^T g(T-t)dt$$

in agreement with our expectations; for if the intensity is so small that the probability of more than one particle arriving within the time of one pulse is negligible, then the probability of arrival of this "first" particle in a pulse is uniform throughout the duration of the pulse; and the energy per pulse is therefore simply mT times the average energy $(1/T)\int_0^T g(T-t)dt$ produced by one particle. For very large α, the situation is quite different: e^{-mt} differs from zero appreciably only near the origin, so that we get

$$E(m)=Ng(T)=Ng_{\max},$$

independent of m, again with agreement of expectation; for if the intensity is so large that a large number of particles are almost certain to arrive within the time of one pulse, then one of them is almost certain to arrive at the very beginning of the pulse, giving rise to current $g_{\max}$ from every pulse. Even though this conclusion is correct, it should be borne in mind that the assumption

[4] Tubes in which photoeffect at the cathode is believed responsible for spread are studied by L. Colli and U. Facchini, Phys. Rev. **88**, 987 (1952), who also give references to earlier work; C. G. Montgomery and D. D. Montgomery, Phys. Rev. **57**, 4030 (1940).

[5] Tubes in which photoeffect in the gas near the wire is believed responsible for spread are studied by F. Alder, E. Baldinger, P. Huber, and F. Metzger, Helv. Phys. Acta **20**, 73 (1947), who also give references to earlier work. E. Greiner, Z. Physik **81**, 543 (1933).

upon which this section is based, and hence also Eq. (6), is not likely to be close to the truth as m gets large.

The integral equation (6) enables us to compute the average output E of the tube as a function of radiation intensity m and the response $g(T-t)$ of the tube to a single particle as a function of time of arrival. For the reverse problem of inferring from an experimentally determined E *vs* m relationship, this equation can be solved[6] for g. We find

$$g(t)=(1/N)\left[E(m,t)+(1/m)\frac{dE(m,t)}{dt}\right] \tag{7}$$

as may be verified by direct substitution and integration by parts.

If one plots several assumed functions $g(T-t)$, e.g.,

$$g_1=k_1(t/T)e^{c_1(t/T)} \quad \text{or} \quad g_2=k_2(t/T)e^{-c_2(t/T)}$$

$$\text{or} \quad g_3=k_3(t/T)^3,$$

with constants k_1, k_2, k_3 so adjusted that $\int_0^T g dt$ is the same in each case, one is struck by their great similarity in shape. Also, response functions E_1, E_2, E_3, pairwise related to g_1, g_2, g_3 by (6) and (7), turn out to be almost indistinguishable graphically in spite of their mathematically different form and the different g functions from which they are derived. In view of this, and the additional fact that the work of this section does not hold at very high intensities, Eq. (7) will be of little use if we wish to calculate g precisely enough from experimental $E(m)$ curves to make inferences concerning the physical processes responsible for discharge spread. This, to be sure, does not detract from the usefulness of Eq. (6), (7) in the practical problem of relating incident radiation intensities and output currents.

Let us conclude the discussion of this approximation with a somewhat more concrete example of a case to which it applies. In Sec. IIIB, we discuss the spread mechanism which may be active in argon and also in halogen filled tubes, and conclude that the conditions for the validity of the approximation of the present section are well satisfied. We are, therefore, justified in applying Eq. (6) to the $g(t)$ found in Sec. IIIB [Eq. (13)]. We thus obtain

$$E(m)/N=m\int_0^T e^{-m(T-t)}e^{kt}/(\beta+e^{kt})dt. \tag{8}$$

This function is plotted in Fig. 1. The physical meaning of β is discussed in Sec. IIIB, and k is so chosen that half the charge flowing in one pulse will pass in one μsec.

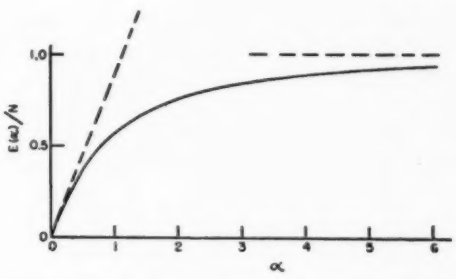


FIG. 1. Example of response of a pulsed tube with spread by photoelectric effect at the cathode as a function of incident radiation intensity, calculated by the "rapid spread" approximation [Eq. (8)]. E/N, the normalized output current is plotted *vs* the incident radiation intensity.

B. Photoelectric Effect at Cathode

We must now briefly consider the physical processes responsible for tube operation. Electrons produced by an incoming particle move towards the anode (the center wire). When they get very close to the wire, they gain enough energy in one mean free path to cause further ionization; thus, an "avalanche" is formed. Each avalanche is believed to be about 10^{-2} cm long. The avalanches give off photons and these produce further avalanches by photoelectric effect, thus causing the spread of the discharge. This photoelectric process may take place in two ways depending on the nature of the tube: either in the gas near the anode wire, or at the cathode wall. We shall discuss the latter process in this Sec. IIIB and the former process in Sec. IIIC.

Assume that it is permissible to apply differential calculus to the problem, i.e., that the discrete events that make up the process occur often enough to permit the process to be treated as a continuous one. At the beginning of a discharge, let the photons produced by one avalanche give rise, by electron production by photoelectric effect at the cathode, to c new avalanches:

$$da=cadt, \tag{9}$$

where $a(t)$ is the number of avalanches at time t. However, on account of the sheath of positive ions, new avalanches can form only on such spots along the wire where none have formed previously; therefore, Eq. (9) will no longer be correct if a substantial part of the wire is already covered by avalanches. Instead we must write

$$da=cap(a)dt, \tag{10}$$

where $p(a)$ is the probability that an electron moving towards the wire will arrive at a place on the wire where no previous avalanche has formed. Since that electron is equally likely to end up at any point on the wire, we may take simply

$$p(a)=(\beta-a)/\beta, \tag{11}$$

where β is the total number of avalanches that will fit in the tube. (If each avalanche is 10^{-2} cm long and the tube is 10 cm long, then $\beta=10^3$. Henceforth, we assume $\beta\gg1$.) Equation (10) thus becomes

$$da=ca(1-a/\beta)dt. \tag{12}$$

Integration gives the number of avalanches at any time t

$$a(t)/\beta=e^{ct}/[(\beta-1)+e^{ct}]. \tag{13}$$

[6] R. V. Churchill, *Operational Mathematics in Engineering* (McGraw-Hill Book Company, Inc., New York, 1944).

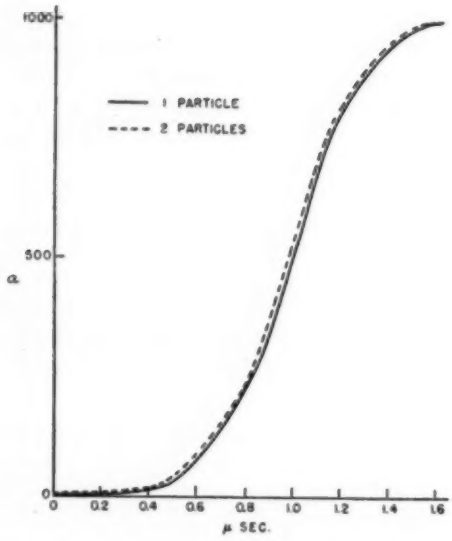


FIG. 2. Example of the response of a tube (not pulsed) with spread by photoelectric effect at the cathode to a single particle, and to two particles. The ordinate is proportional to the total charge flowing through the tube before time t.

Solid line: response to a single particle that came in at time 0 [Eq. (13)]. Dashed line: response to two particles, of which the first came in at time 0, the second sometime before time t, averaged over the possible arrival times of the second particle [Eq. (17)].

Here $(\beta-1)$ is the correct integration constant to give $a(0)=1$. Observe further that (13) leads to $a(\infty)=\beta$, which is physically necessary and is a statement of well-known fact that in a Geiger tube operated on a constant electrode voltage all pulses reach the same height, and to an exponential rise in $a(t)$ for small t. Equation (13) is plotted in Fig. 2 (heavy line). In Fig. 2 we have taken $\beta=1000$ and have fixed c to have the charge reach half of its final value in 10^{-6} sec.

So far we have considered only the multiplication of one avalanche. In this paper, we are primarily concerned with the effect of several particles (several separate initial avalanches). In that case (12) no longer holds for the instants $t=t_1$, $t=t_2$, $\cdots$, $t=t_n$ that the n particles come in. (Assume $0<t_1<t_2<\cdots<t_n$.) Formally we may write, correctly for all t

$$da=\left[ca\,p(a)+\sum_l^n p(a)\delta(t-t_l)\right]dt, \tag{14}$$

$\delta(x)$ being the Dirac delta. At all times not including any t_l, Eq. (14) is identical with (13) and may be integrated in the same way. The second term in (14) tells us to match up the resulting piecewise solutions by

$$a(t_l+0)=a(t_l-0)+p(a)$$

or, with $\beta\gg1$ by

$$a(t_l+0)=a(t_l-0)+1.$$

We thus find

$$a_n(t)/\beta=e^{ct}/[d_n+e^{ct}], \tag{15}$$

where the constant d_n is now a function of $t_1, t_2, \cdots, t_k$, given by the recursion formula

$$d_n=\frac{(\beta-1)d_{n-1}-e^{ct_n}}{\beta+1+e^{-ct_n}d_{n-1}}.$$

$$d_1=(\beta-1)e^{ct_1}$$

We have written a_n to indicate that this is the charge obtained from n particles. If we have only one particle, $a_1(t)$ of Eq. (15) is seen to go into Eq. (13), as it should. To get the average charge $\langle a_n\rangle$ from a pulse resulting from n particles, we average over the possible arrival times which are

$$t_{n-1}<t_n<T \quad \text{for} \quad t_n$$
$$t_{l-1}<t_l<t_{l+1} \quad \text{for} \quad t_l \quad \text{for} \quad 1<l<n$$
$$0<t_1<T \quad \text{for} \quad t_1,$$

and get

$$\langle A_n\rangle=\int_0^T I_1dt_1\int_{t_1}^T I_2dt_2\int_{t_2}^T I_3dt_3 \times\int_{t_3}^T\cdots\int_{t_{n-2}}^T I_{n-1}dt_{n-1}\int_{t_{n-1}}^T I_ndt_n \tag{16}$$

with $I_k\equiv g(k,t_k;n,T)$ as previously defined. Only the simplest of the multiple integrals (16) are easily obtained in closed form. As an illustration, we have evaluated $J(t)\equiv(1/t)\int_0^t a_2dt_2$ with $t_1=0$; this is the number of avalanches at time t due to two particles of which the first came in at $t=0$, averaged over the arrival time of the second. We obtain

$$J(t)=\beta\left\{1-\frac{\beta}{\beta+e^{ct}}\left[1+\frac{1}{ct}\ln\frac{\beta+1+e^{ct}}{\beta+2e^{ct}}\right]\right\}. \tag{17}$$

Note $J(0)=2$, $J(\infty)=\beta$. Figure 2, where the function $J(t)$ is plotted (dashed line) with β and C taken as before shows it to be almost undistinguishable from a_1. This means that the pulse height is almost completely determined by the time at which the first particle comes in; or, that the assumption of Sec. IIIA applies well to the model considered in the present section. This is not hard to understand qualitatively. The first particle multiplies rapidly (exponentially) at first, and less rapidly later when it has increased to a substantial fraction of the maximum height that the pulse is going to attain. If the second particle comes in substantially later than the first (and this is what is likely to happen with low and medium incident radiation intensities), then the first particle will have given rise to so many avalanches by then that the addition of one extra one from the second particle will mean no appreciable increase. If, on the other hand, the second particle comes in immediately after the first one [the extreme case here is simultaneous arrival, forcing us to compare $a_1(0)=1$ with

$a_2(0)=2$] then the immediate discrepancy between the two curves is large; but this is not likely to happen except with very large incident radiation intensities, and in that case, it is likely to happen early in the pulse so that the current will reach its maximum value anyhow. The effect of the second particle, and by the same reasoning, of all other particles after the first, is thus likely to be small whatever the time of its arrival, and the assumption Sec. IIIA that only the first particle's effect need be considered is likely to be a good one for this model with all except very high intensities (e.g., more than about 5 particles per pulse).

C. Linear Spread

In this section we consider tubes in which avalanches reproduce by means of photoelectric effect in the gas filling. The photoelectric process, and the resulting daughter avalanches, is thus likely to occur near the wire close to the location of the parent avalanche, so that the discharge will seem to spread along the wire with a definite average velocity, which we shall call v. The total charge per pulse, being, as before, proportional to a total number of avalanches, or, in this case, to the total length along the wire over which the discharge has spread, will therefore depend on both the time t_1 of arrival and the position x_1 of arrival in the tube of the particle, even if only one particle arrives. E.g., if the tube voltage is on $0<t<T$ and the particle arrives at $t=t_1$ in a tube of length $2L$, then the output pulse is either $2v(T-t)$ or $2L$, whichever is smaller, if it arrives in the center of the tube, but only $v(T-t)$ or $2L$, whichever is smaller, if it arrives at the very end of the tube, because in the latter case the discharge can spread in one direction only. More generally, to find F_k we must find how much of the length of the tube along the wire is going to be covered at least once by the discharges resulting from the k particles which arrive, respectively, at x_1t_1, at x_2t_2, $\cdots$, at x_kt_k, and then average over all possible $x_1, \cdots, x_k, t_1, \cdots, t_k$. We can show [Appendix I, Eq. (A6)] that

$$F_k/A=1-\frac{1}{L}\int_0^L [1-q_v]^k dx, \tag{18}$$

where $q_v(x)$, defined in Appendix I, is the probability that the point $x(-L<x<L)$ will be covered by the discharge spreading at speed v resulting from the spread of a single particle arriving somewhere sometime in an interval of T seconds, and A is the maximum possible value of F_k, attained when the entire tube is covered by the discharge. Note that formally $F_0=0$. We can thus write Eq. (3) as

$$E/N=\sum_{k=0}^{\infty}\frac{\alpha^k e^{-\alpha}}{k!}F_k, \tag{19}$$

and by putting (18) into (19) and summing the series over k obtain

$$E/NA=1-\frac{1}{L}\int_0^L e^{-\alpha q_v(x)}dx. \tag{20}$$

q being positive definite, we see from this form that E is an increasing function of α, that $E(0)=0$, and that $E(\infty)/NA=1$. This is of course required physically. $q_v(x)$ is evaluated in Appendix II. Equation (20) then gives

$$E/NA=1-\Psi(\alpha), \tag{21}$$

where

$$\Psi(\alpha)=\begin{cases} e^{-c\alpha/2}\{1-c+2(c/\alpha)^{\frac{1}{2}}\Phi[\frac{1}{2}(\alpha c)^{\frac{1}{2}}]\} & 0<c<1 \\ (2c/\alpha)^{\frac{1}{2}}\{e^{-(1-c/2)\alpha}\Phi\{[(c/2)^{\frac{1}{2}}-(1/2c)^{\frac{1}{2}}]\alpha^{\frac{1}{2}}\}+\sqrt{2}e^{-c\alpha/2}\{\Phi[\frac{1}{2}(\alpha c)^{\frac{1}{2}}]-\Phi[(\sqrt{c}-\sqrt{1/c})\alpha^{\frac{1}{2}}\} & 1<c<2 \\ (2c/\alpha)^{\frac{1}{2}}e^{-(1-c/2)\alpha}\Phi[(\alpha/2c)^{\frac{1}{2}}] & 2<c<\infty. \end{cases}$$

$c=vT/L$ is a constant depending on the tube length and the duration of the applied pulse, and $\Phi(x)=\int_0^x e^{+t^2}dt$ is a known, tabulated[7] function. In Fig. 3, $E(\alpha)$ is plotted for several values of the parameter c. The cases of interest are those for which c is of the order of unity. For $c\to 0$ and $c\to\infty$, (21) reduces to expressions which can be found from much simpler considerations. Thus, for $T\to 0$ (or $c\to 0$), Eq. (21) approaches

$$E/AN=\tfrac{1}{2}\alpha c\left[(1-c/6)+\tfrac{1}{4}\alpha c\left(1-\frac{17}{60}c\right)+\cdots\right]. \tag{22}$$

For physical reasons in this case, one would expect that at most one particle, giving, on the average, rise to one-half of the possible maximum charge per pulse, would come in in one pulse, giving rise to a charge of $(\frac{1}{2})vT/L$, in agreement with the first term in Eq. (22). For $T\to\infty$ Eq. (21) approaches

$$E/NA=1-e^{-\alpha}. \tag{23}$$

This, in turn, is identical with the well-known result obtained by assuming that every voltage pulse contributes the maximum charge if one or more particles come in. (If the probability that one ball will be in a certain box is T, then the probability of at least one out of αT balls being in the box is $1-(1-T)^{\alpha T}$ or, in the appropriate limit, $1-e^{-\alpha}$.)

IV. CONCLUSIONS

The advantage of pulsed operation is that the range of usefulness of the Geiger tube is thereby extended greatly into the higher radiation intensities; the shorter the applied pulse, the greater is this extension. The drawback associated with such operation is that, since particles arriving near the end of an applied pulse do

[7] E. Jahnke and F. Emde, *Tables of Functions* (Dover Publications, New York, 1945), p. 32.

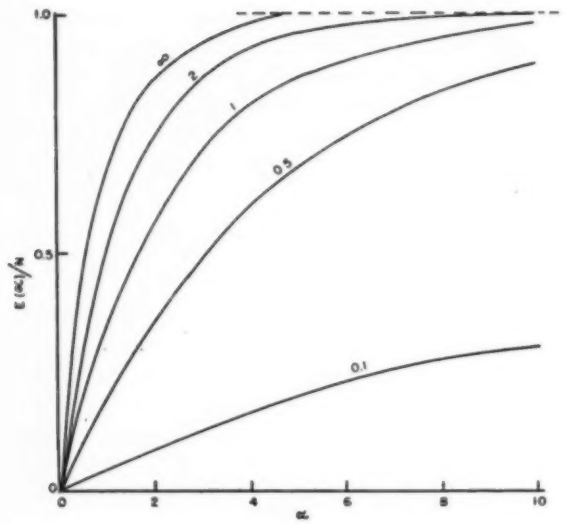


FIG. 3. Response of a pulsed tube with spread by photoelectric effect in the gas as a function of incident radiation intensity, for various ratios of spread velocity to tube length. E/N, the normalized output current, is plotted *vs* the incident radiation intensity. The number labeling the various curves are values of the parameter $c=vT/L$.

not have time to grow to their full height, the simple linear relation between output current and radiation intensity is lost; the shorter the applied pulse, the greater are these nonlinear effects. This paper is an attempt to study these nonlinear effects, and to obtain, analytically if possible, the relation between radiation intensity and output current under pulsed operation. It is found that satisfactory approximate expressions, valid for all but extremely high radiation intensities, can be obtained independent of the mechanism of discharge spread if only the spread is fast initially. Spread by photoelectric effect at the cathode is an example of a mechanism satisfying this condition; its detailed treatment without recourse to the above approximation leads to multiple integrals so cumbersome as to be of little use for practical computations. If, on the other hand, spread is caused by photoelectric effect in the gas near the wire, the desired relationship can be obtained rather simply in closed form.

This work was begun at the suggestion of Mr. S. W. Lichtman. I wish to thank Mr. Lichtman, Dr. H. Friedman, and Dr. Talbot Chubb for several useful discussions. I am particularly grateful to Dr. Arthur Linz, Jr., for a critical reading of the manuscript.

APPENDIX I

Consider the following problem in geometrical probabilities.

k needles (one-dimensional objects), of length $2\tau_1$, $2\tau_2, \cdots, 2\tau_k$, respectively, are tossed at random unto the x axis in such a fashion that their midpoints fall on $x_1, x_2, \cdots x_k$, respectively, where $-L<x_n<L$ for all n. All points of arrival $-L<x<L$ for midpoints are assumed equally likely, and all needles fall parallel to the x axis. Find what fraction of the length $2L$ between $-L$ and L will, on the average, be covered at least once if $\tau_1, \tau_2, \cdots, \tau_k$ are chosen at random subject only to $\tau_n<M$ for all n.

This problem is seen to be equivalent to the one of finding F_k in Sec. III if we make the following identifications: length of tube is $2L$, $vt_n=\tau_n$, $vT=M$. We use the formulation of the preceding paragraph only because it simplifies the language slightly. The situation is illustrated in Fig. 4.

Let Λ_k be the length $-L<x<L$ that is covered at least once when a total of k needles are used. In case $k=1$ (one needle of length $2\tau_1$), the length Λ_1 of $-L<x<L$ that is covered at least once is simply the length $D^{(1)}(x_1,\tau_1)$ of needle 1 that overlaps with $-L<x<L$:

$$\Lambda_1=D^{(1)}(x_1,\tau_1).$$

If we have $k=2$, we have

$$\Lambda_2=D^{(1)}(x_1,\tau_1)+D^{(1)}(x_2,\tau_2)-D^{(2)}(x_1,\tau_1,x_2,\tau_2),$$

where $D^{(2)}$ is the length on $-L<x<L$ covered simultaneously by both of the two needles (the length over which the two needles overlap). If $k=3$, then we have

$$\Lambda_3=D^{(1)}(1)+D^{(1)}(2)+D^{(1)}(3)-D^{(2)}(1,2)-D^{(2)}(1,3)-D^{(2)}(2,3)+D(1,2,3),$$

where the arguments (x_m,τ_m) have been abbreviated (m) and $D^{(3)}(1,2,3)$ is the length covered simultaneously by all three needles (the length over which all three needles overlap). Generally, we see that

$$\Lambda_k=\sum_{n=1}^{k} D^{(1)}(n)-\sum_{m>n}^{k} D^{(2)}(n,m)+\sum_{m>n>l}^{k} D^{(3)}(l,n,m)\cdots+(-)^{k-1}D^{(k)}(1,2,\cdots k)$$
$$=\sigma_{1k}-\sigma_{2k}+\sigma_{3k}\cdots+(-)^{k-1}\sigma_{kk}. \tag{A1}$$

What we want is the average of Λ_k,

$$F_k=\frac{1}{(2L)^k}\int_{-L}^{L}\int_{-L}^{L}\cdots\int_{-L}^{L}\frac{1}{M^k}\int_0^M\int_0^M\cdots\times\int_0^M\Lambda_k d\tau_1\cdots d\tau_k dx_1\cdots dx_k,$$

which we abbreviate

$$F_k=I_M(\tau)I_L(x)\Lambda_k.$$

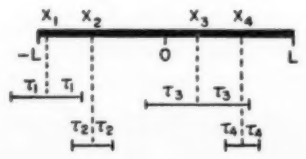


FIG. 4. Sketch of the linear spread in a tube in which four particles have arrived at different times and locations. See Appendix I.

Carry out this integration on the right-hand side of (A1). Because the variables integrated over appear symmetrically, we have

$$I_M(\tau)I_L(x)D^{(1)}(n)=I_M(\tau)I_L(x)D^{(1)}(m)$$

for all m and n;

$$I_M(\tau)I_L(x)D^{2}(m,n)=I_M(\tau)I_L(x)D^{(2)}(k,l)$$

for all m, n, k, and l; etc.

We can then write

$$F_k=\alpha_{1k}0^{(1)}-\alpha_{2k}0^{(2)}+\alpha_{3k}0^{(3)}\cdots+(-)^{k-1}0^{(k)}, \quad \text{(A2)}$$

where we have written

$$0^{(n)}\equiv I_M(\tau)I_L(x)D^{(n)} \quad \text{(A3)}$$

and where α_{nk} is just the number of terms in σ_{nk}. The physical meaning of $0^{(n)}$ is the average length covered n times by n needles. By direct counting we find

$$\alpha_{nk}=k!/n!(n-k)!\equiv\binom{k}{n}.$$

(A2) then becomes

$$F_k=\sum_{n=1}^{k}\binom{k}{n}(-)^{n+1}0^{(n)},$$

or, if we stipulate $0^{(0)}=1$,

$$F_k=1-\sum_{n=0}^{k}(-)^n\binom{k}{n}0^{(n)}. \quad \text{(A4)}$$

Let us then proceed to calculate the $0^{(n)}$, using the interpretation given two sentences up.

Let $p(x,\tau_1)$ be the probability that the point x be covered by the needle of length $2\tau_1$ which has been tossed randomly at the x axis in the prescribed fashion. Then $p(x,\tau_1)p(x,\tau_2)$ is the probability that the point x will be covered by both needle 1 of length $2\tau_1$ and needle 2 of length $2\tau_2$, and $p(x,\tau_1)p(x,\tau_2)\cdots p(x,\tau_k)$ is the probability that x will be covered by needles of length $2\tau_1, 2\tau_2, \cdots, 2\tau_k$ together. The probable total length covered k times is obtained by summing this over x, and the average of this quantity, independent of the length of the needles, is obtained by averaging over the τ_k. Thus

$$\begin{aligned}0^{(k)}&=\frac{1}{M^k}\int_0^M\cdots\int_0^M\frac{1}{2L}\int_{-L}^{L}p(x\tau_1)p(x\tau_2)\cdots\\&\qquad\times p(x\tau_k)dxd\tau_1\cdots d\tau_k\\&=\frac{1}{2L}\int_{-L}^{L}\left[\frac{1}{M}\int_0^M p(x\tau_1)d\tau_1\right]^k dx\\&=\frac{1}{2L}\int_{-L}^{L}(q_M(x))^k dx\\&=\frac{1}{L}\int_0^L(q_M(x))^k dx,\end{aligned} \quad \text{(A5)}$$

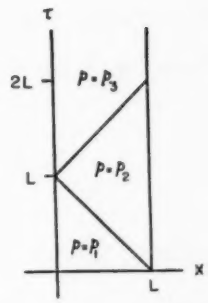


FIG. 5. Behavior of $p=p(x,\tau)$ in the $x-\tau$ plane, as given by Eq. (A7). We have abbreviated $p_1=\tau/L$, $p_2=(\tau+L-x)/2L$, $p_3=1$.

where

$$q_M(x)=\frac{1}{M}\int_0^M p(x\tau)d\tau.$$

The last step follows from the physical requirement that p, and therefore also q, be an even function of x. Note that, formally, $0^{(0)}=1$ follows from (A5) automatically. If we put (A5) into (A4), and sum the resulting series over n, we get

$$F_k=1-\frac{1}{L}\int_0^L[1-q_M(x)]^k dx. \quad \text{(A6)}$$

The evaluation of $q_M(x)$ is indicated in Appendix II.

APPENDIX II

$p(x,\tau)$ has been defined in Appendix I as the probability that point x on $-L<x<L$ will be covered by the needle of length 2τ. That is,

$$p(x,\tau)=\frac{\text{positions of needle of length } 2\tau \text{ for which } x \text{ is covered}}{\text{totality of possible positions of needle of length } 2\tau}.$$

With this, we see from Fig. 4 directly that

for $0<\tau<L$

$$p(x,\tau)=\begin{cases}2\tau/2L & 0<x<L-\tau\\(\tau+L-x)/2L & L-\tau<x<L\end{cases}$$

for $L<\tau<2L$ (A7)

$$p(x,\tau)=\begin{cases}2L/2L & 0<x<\tau-L\\(\tau+L-x)/2L & \tau-L<x<L\end{cases}$$

for $2L<\tau<\infty$

$$p(x,\tau)=2L/2L \qquad \text{all } 0<x<L.$$

The situation is summarized graphically in Fig. 5.

From Fig. 5, we see that

$$q_M(x)=\frac{1}{M}\int_0^M p(x,\tau)d\tau$$

is explicitly given by the following:

if $0<M<L$

$$Mq_M(x)=\begin{cases}\displaystyle\int_0^M p_1 d\tau & 0<x<L-M\\ \displaystyle\int_0^{L-x} p_1 d\tau+\int_{L-x}^M p_2 d\tau & L-M<x<L\end{cases}$$

if $L<M<2L$

$$Mq_M(x)=\begin{cases}\displaystyle\int_0^{L-x} p_1 d\tau+\int_{L-x}^{L+x} p_2 d\tau+\int_{L+x}^M p_3 d\tau & 0<x<M-L\\ \displaystyle\int_0^{L-x} p_1 d\tau+\int_{L-x}^M p_2 d\tau & M-L<x<L\end{cases}$$

if $2L<M<\infty$

$$Mq_M(x)=\int_0^{L-x} p_1 d\tau+\int_{L-x}^{L+x} p_2 d\tau+\int_{L+x}^M p_3 d\tau \qquad \text{all } 0<x<L.$$

These integrals are all elementary. In the body of the paper, we write $q_r(x)$ instead of $q_M(x)$.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 3 MARCH, 1954

Potential Distribution and Prevention of a Space-Charge-Induced Minimum Between a Plane Secondary Electron Emitter and Parallel Control Grid*†

GEORGE C. SPONSLER‡
Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received June 3, 1953)

Formulas are developed which define the profile of the electrostatic potential distribution in the semi-infinite, space-charge-filled region between a plane emitter of secondary electrons and a parallel retarding-field control grid. The calculation assumes a known secondary-electron energy distribution and a cosine angular distribution. In practical application, the formulas must be evaluated numerically or graphically since the energy distribution cannot be expressed analytically except over limited regions. A method of determining the critical spacing between electrodes which will prevent formation of a space-charge-induced minimum in the potential distribution between the emitter and control grid is presented.

THIS problem arose during the development of an electron tube in which a control grid was employed to limit the current of secondary electrons emitted by a target electrode which was bombarded by a primary electron beam of uniform current density. The plane target and parallel grid were mounted perpendicularly to the primary electron beam which passed through the grid before impinging on the target. In order that the secondary-electron current be controlled by the applied grid potential in a linear manner, it was necessary that the effect of space charge be minimized to the point where there would be no space-charge-induced minimum or "well" in the electrostatic potential distribution between the target and the control grid. (This latter electrode was operated at a negative potential difference with respect to the target; thus, the secondary electrons were subjected to a retarding electric field.) It was recognized that the space-charge effect could be minimized by maintaining a sufficiently small spacing between target and grid, thereby increasing the electric field strength. To estimate the required spacing, it is necessary to determine the distribution of the electric potential in the region between the two electrodes. This potential distribution is derived in the following analysis, and the critical electrode spacing needed to prevent formation of a space-charge-induced potential minimum is estimated.

In order to facilitate the analysis, certain assumptions were made. First, it was assumed that the problem was one-dimensional, and that the various charge and current densities were uniform everywhere on a particular cross section in the semi-infinite region between the plane target and parallel grid. The primary electron beam was assumed to pass through the control grid which intercepted no primary electrons. All secondary

* Part of a dissertation submitted to the faculty of Princeton University for the degree of Doctor of Philosophy.

† The research reported in this document was supported jointly by the Army, Navy, and Air Force under contract with the Massachusetts Institute of Technology, and also in conjunction with the RCA Laboratories at Princeton, New Jersey.

‡ Staff member, Lincoln Laboratory, Massachusetts Institute of Technology.

electrons were considered to be collected or removed from the region as soon as they reached the control grid (or target, if totally reflected).

Finally, in treating the energy distribution of the secondary electrons, it was assumed that $J(eV)$[1] is proportional to the experimental $i(eV)$[2] (see Fig. 1), but independent of the direction of emission, and that the angular distribution was a cosine function.[3]

The secondary-electron speed distribution would be given by

$$J(v)dv=J(eV)dV,$$

or as

$$eV=\tfrac{1}{2}mv^2,$$

$$J(v)=\left.\frac{J(eV)}{(e/2mV)^{\frac{1}{2}}}\right|_{eV=\frac{1}{2}mv^2}. \tag{1}$$

The secondary electrons leave the target with an angular distribution taken to be a cosine function. The number of secondary electrons per second which are emitted by a differential element of area $d\sigma$ into a cone bounded by θ and $\theta+d\theta$ (cf., Fig. 2) would be proportional to

$$J(v)\sin\theta\cos\theta dvd\theta d\sigma.$$

This same group of electrons can be characterized by the distribution of their normal (x directed) velocity component, whence

$$J[v_x(\theta)]=\left.\frac{J(v)}{\cos\theta}\right|_{v=v_x/\cos\theta}. \tag{2}$$

For practical applications, the current densities need to be normalized by evaluating

$$N\int_0^{\pi/2}\int_0^{v_{pri}} J(v)\sin\theta\cos\theta dvd\theta d\sigma=\delta J_{pri}d\sigma \tag{3}$$

for N. Hereafter, we shall assume that the normalization constant is implicitly included in the various current density functions.

The total number of secondary electrons leaving $d\sigma$ per second possessing the normal velocity component between v_x and v_x+dv_x is

$$I_\sigma=\int_0^{\pi/2}\left\{J(v)\Big|_{v=v_x/\cos\theta}\sin\theta dv_x d\sigma\right\}d\theta;$$

hence, the associated total current density normal-velocity-component distribution becomes

$$J_\tau(v_x)=\int_0^{\pi/2}J(v)\Big|_{v=v_x/\cos\theta}\sin\theta d\theta. \tag{4}$$

[1] See glossary at end of paper for meaning of symbols not defined in the text.

[2] E. Rudberg, Phys. Rev. **50**, 120 (1936).

[3] J. L. H. Jonker, "On the Theory of Secondary Electron Emission," Phillips Research Reports (February 7, 1952), pp. 1–20.

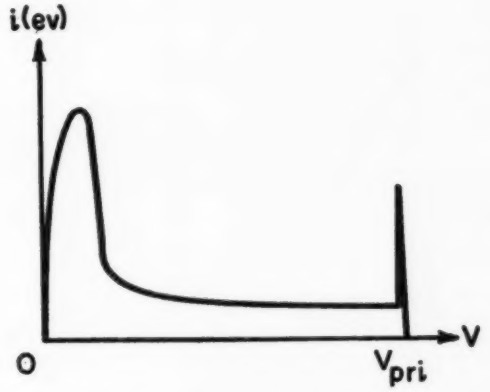


FIG. 1. Typical secondary-electron energy distribution.

Having now characterized the total population of secondary electrons by the distribution of their normal velocity components, we shall replace the actual angularly distributed secondary-electron population by an equivalent, hypothetical one possessing the same normal velocity component alone. Assuming that the real electrons in the interelectrode region are distributed uniformly in cross sections perpendicular to the x axis, this population substitution is permissible as the electric field has no component perpendicular to the x axis. This equivalent secondary-electron population has the same net current density and spatial charge distribution as the actual population, but its electrons move only in the normal (or x) direction.

The equivalent population possesses a current-density energy distribution given by[4]

$$J_\tau(eV)=\left.\frac{J_\tau(v_x)}{(2mV/e)^{\frac{1}{2}}}\right|_{v_x=(2eV/m)^{\frac{1}{2}}}. \tag{5}$$

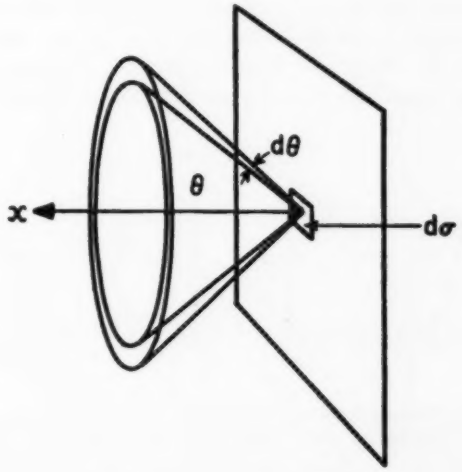
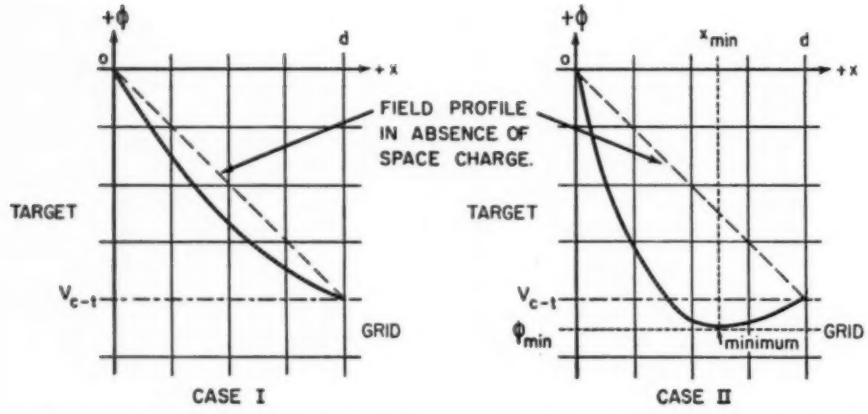


FIG. 2. Scattering geometry.

[4] The differential multiplier of $J(eV)$ is actually $d(eV)=edV$. The coefficient e is considered to be included with the intrinsic normalization constant.



No Potential Minimum Exists Between Electrodes

Potential Minimum Exists Between Electrodes

FIG. 3. Electrostatic potential profiles.

This equivalent energy distribution is employed in the subsequent analysis first to determine the charge density and secondly, the electric potential distribution in the region between grid and target.

It is convenient to distinguish between two types of potential distributions which are portrayed in Fig. 3. Case I exists when the effect of the space charge is not great enough to lower the electric field potential below the effective control grid bias voltage V_{c-t}.[5] Case II exists when the space charge is quite large and, as a result, a potential minimum exists between the two electrodes. In both cases, the target is taken to be the potential reference point and is positive with respect to the grid. In the absence of space charge, the electrostatic field potential would be as depicted by the dashed straight lines. The effect produced by the presence of a space charge caused by the electrons is to lower the potential profile below the dashed line.

Certain of the secondary electrons will possess energy sufficient to permit them to pass "over" the minimum potential between the grid and the target. This minimum may be either the effective grid bias voltage V_{c-t} as in Case I, or the minimum caused by the space charge in Case II. Other secondary electrons will possess insufficient energy to pass the potential minimum and will be reflected back toward the target. These two types of secondary electrons, plus the electrons of the primary electron beam, constitute the space charge.

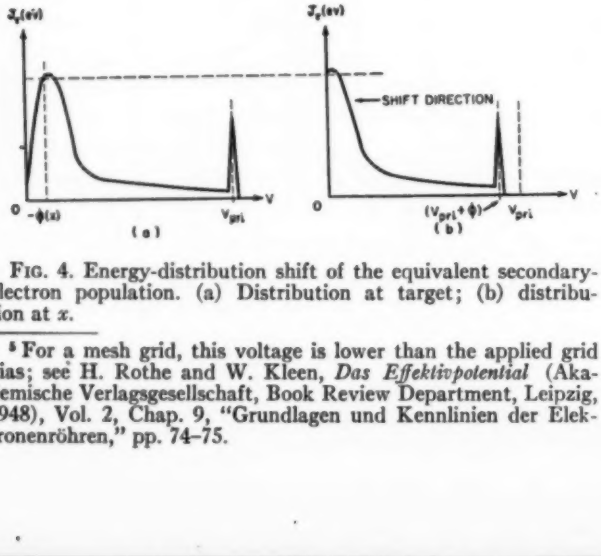


FIG. 4. Energy-distribution shift of the equivalent secondary-electron population. (a) Distribution at target; (b) distribution at x.

The hypothetical electrons moving away from the target (the "outgoing" secondary electrons) that pass a plane at potential ϕ and distance x from the target will have an energy distribution the shape of which is the same as that of the equivalent secondary-electron population at the target surface but shifted down the energy (or the equivalent-voltage) axis by an amount equal to the potential difference of the electric field between that plane and the target. All portions of this "shifted" distribution that would lie in the region of negative energies as a result of the shift are, of course, nonexistent, since this region corresponds to the portion of the secondary-electron current that has been reflected back toward the target. Figures 4(a) and 4(b) illustrate this argument.

The time of flight required by the secondary electrons to reach the plane at x does not alter the energy distribution at x, once an equilibrium state has been reached. If the currents and fields are time-variable, it is only required that the frequency have a period long in comparison with the longest electron transit time.

The net secondary-electron current passing the control grid is the resultant current of the outgoing and reflected secondary electrons. The charge density will be different for planes at different distances from the target. In order to compute the electric field potential distribution as a function of x, it is first necessary to determine the charge density throughout the inter-electrode region as a function of the potential, ϕ.

Continuity of current demands that, for the small group of electrons[6] of velocity between v and $v+dv$, $dJ=-vd\rho$, or

$$d\rho=-\frac{dJ}{(2eV/m)^{\frac{1}{2}}}. \tag{6}$$

[5] For a mesh grid, this voltage is lower than the applied grid bias; see H. Rothe and W. Kleen, *Das Effektivpotential* (Akademische Verlagsgesellschaft, Book Review Department, Leipzig, 1948), Vol. 2, Chap. 9, "Grundlagen und Kennlinien der Elektronenröhren," pp. 74–75.

[6] Direct reference to the equivalent electron population will hereafter be dropped, but all statements will be understood to refer to this substituted, hypothetical population.

This equation holds, of course, only for $V>\phi$ at the particular x.

Using Eq. (6), the contributions to the charge density made by the outgoing and reflected secondary electrons are:

$$\rho_+=-\left(\frac{m}{2e}\right)^{\frac{1}{2}}\int_{-\phi}^{V_{\rm pri}}\frac{J_\tau(eV)}{(V+\phi)^{\frac{1}{2}}}dV,$$

and

$$\rho_-=-\left(\frac{m}{2e}\right)^{\frac{1}{2}}\int_{-\phi}^{-\phi_{\rm min}}\frac{J_\tau(eV)}{(V+\phi)^{\frac{1}{2}}}dV,$$

since $dJ=J_\tau(eV)dV$.

Note that $\phi(x)$, $\phi_{\rm min}$ and V_{c-t} are negative (while $V_{\rm pri}$ and V are positive); hence, $-\phi$ is positive. The portions of the initial-charge-density energy distribution which correspond to the outgoing and reflected charge densities at x are portrayed by the shaded areas of Figs. 5(a) and 5(b).

The upper limit of the various integrals cannot be greater than $V_{\rm pri}$ as none of the secondary electrons possesses an energy greater than that of the primary beam.

We consider the primary-electron beam to be energetically monochromatic; hence, the contribution to the total charge density made by it is given by

$$\rho_{\rm pri}=-\left(\frac{m}{2e}\right)^{\frac{1}{2}}\frac{J_{\rm pri}}{\sqrt{V_{\rm pri}}},\quad V_{\rm pri}\gg|V_{c-t}|.$$

The total charge density is the sum of these various contributions:

$$\rho_t(\phi)=-\left(\frac{m}{2e}\right)^{\frac{1}{2}}\left\{\frac{J_{\rm pri}}{\sqrt{V_{\rm pri}}}+\int_{-\phi_{\rm min}}^{V_{\rm pri}}\frac{J_\tau(eV)}{(V+\phi)^{\frac{1}{2}}}dV+2\int_{-\phi}^{-\phi_{\rm min}}\frac{J_\tau(eV)}{(V+\phi)^{\frac{1}{2}}}dV\right\}. \quad (7)$$

In this equation, part of the ρ_+ integral has been added to the ρ_- integral which accounts for the factor 2 of the last term; ϕ and $\phi_{\rm min}$ are as yet unknown.

We must here differentiate between Case I- and Case II-type fields. For Case I, $\phi_{\rm min}$ of the preceding analysis is the effective grid voltage V_{c-t}. For Case II, $\phi_{\rm min}$ is the space-charge-induced minimum potential existent between the electrodes at a distance $x_{\rm min}$ from the target (see Fig. 3). To treat Case II, two sets of calculations must be made: one for the region $0\leqslant x\leqslant x_{\rm min}$, and the other for the region $x_{\rm min}\leqslant x\leqslant d$, where d is the electrode spacing. These two solutions must then be matched at $x=x_{\rm min}$ to give the complete Case II solution. For the region $0\leqslant x\leqslant x_{\rm min}$, the Case II calculation follows along exactly like that of Case I. Calculations for the other region of Case II are similar, except there are no reflected secondary electrons ($\rho_-=0$).

To determine $\phi(x)$, we must solve the Poisson equation, which, in *mks* units, becomes

$$\frac{d^2\phi(x)}{dx^2}=-\frac{\rho_t(\phi)}{\epsilon}. \quad (8)$$

In the present instance, $\rho_t=\rho_t(\phi)$.

Substituting $q=d\phi/dx$, the Poisson equation becomes

$$q\frac{dq}{d\phi}=-\frac{\rho_t(\phi)}{\epsilon}, \quad (9)$$

where ρ_t must be inserted from Eq. (7). The respective integrals of this equation are:

CASE I

$$\int_{q_d}^{q}qdq=-\frac{1}{\epsilon}\int_{V_{c-t}}^{\phi}-\rho_t(\phi')d\phi',$$

as

$$q_d=\left.\frac{d\phi}{dx}\right|_{x=d}\quad\text{and}\quad\phi=V_{c-t}\quad\text{at}\quad x=d.$$

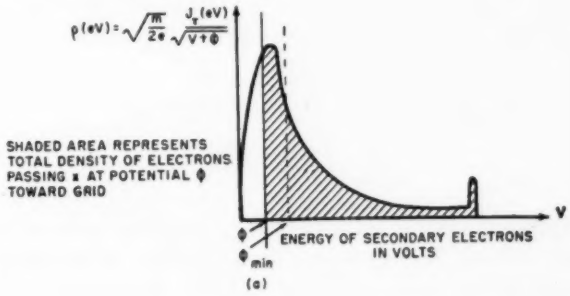
Integrating and resubstituting $q=d\phi/dx$,

$$\frac{d\phi}{dx}=-\left(\frac{2}{\epsilon}\int_{V_{c-t}}^{\phi}-\rho_t(\phi')d\phi'+q_d^2\right)^{\frac{1}{2}}. \quad (10)$$

CASE II, REGION $0\leqslant x\leqslant x_{\rm min}$

$$\int_0^q qdq=-\frac{1}{\epsilon}\int_{\phi_{\rm min}}^{\phi}-\rho_t(\phi')d\phi',$$

as $q=0$ when $\phi=\phi_{\rm min}$.


SHADED AREA REPRESENTS TOTAL DENSITY OF ELECTRONS PASSING x AT POTENTIAL φ TOWARD GRID
V
ENERGY OF SECONDARY ELECTRONS IN VOLTS
φ min
(a)

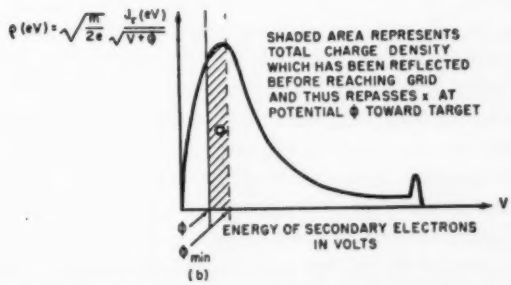

SHADED AREA REPRESENTS TOTAL CHARGE DENSITY WHICH HAS BEEN REFLECTED BEFORE REACHING GRID AND THUS REPASSES x AT POTENTIAL φ TOWARD TARGET
V
ENERGY OF SECONDARY ELECTRONS IN VOLTS
φ min
(b)

FIG. 5. Portions of the initial-charge-density energy distribution corresponding to outgoing and reflected electrons at distance x and potential ϕ. (a) Outgoing electrons; (b) reflected electrons.

Integrating and resubstituting,

$$\frac{d\phi}{dx}=-\left(\frac{2}{\epsilon}\int_{\phi_{\min}}^{\phi}-\rho_t(\phi')d\phi'\right)^{\frac{1}{2}}. \tag{11}$$

In these equations, q_d and $\phi_{\min}$ are constant parameters, as yet unknown, to be determined by subsequent steps of the analysis. The minus sign is chosen before the square root as the gradients are negative in the respective regions.

The terms under the respective square roots are explicit functions of ϕ alone; hence, Eqs. (10) and (11) are separable and yield x as a function of ϕ:

CASE I

$$x=\int_0^{\phi}-\frac{d\phi''}{\left(\frac{2}{\epsilon}\int_{V_{c-t}}^{\phi''}-\rho_t(\phi')d\phi'+q_d{}^2\right)^{\frac{1}{2}}}. \tag{12}$$

CASE II, REGION $0\leqslant x\leqslant x_{\min}$

$$x=\int_0^{\phi}-\frac{d\phi''}{\left(\frac{2}{\epsilon}\int_{\phi_{\min}}^{\phi}-\rho_t(\phi')d\phi'\right)^{\frac{1}{2}}}, \tag{13}$$

as, for both cases, when $x=0$, $\phi=0$; $\rho_t(\phi)$ is given by Eq. (7). Both these quadratures yield positive values for x.

In Case I, we determine q_d by setting x equal to the screen-target spacing d, set $\phi=V_{c-t}$, and solve the resulting equation for q_d; namely,

$$d=\int_0^{V_{c-t}}\frac{d\phi''}{-\left(\frac{2}{\epsilon}\int_{V_{c-t}}^{\phi''}-\rho_t(\phi')d\phi'+q_d{}^2\right)^{\frac{1}{2}}}. \tag{14}$$

Although simple in theory, the foregoing analysis is most difficult to follow in practical cases where most computation must be done either numerically or graphically.

Calculation of the limiting distribution between a Case I and a Case II field yields the critical electrode spacing required to prevent formation by the space charge of a potential minimum. This limiting case is determined from Eq. (12) by letting the field gradient q_d equal zero when $x=d_0$, where d_0 is the critical grid-target spacing sought. The spacing d_0 should be calculated for the smallest effective grid potential $|V_{c-t}|_{\min}$ anticipated. This limiting spacing, as so calculated, represents the largest spacing for which there will be no space-charge-induced potential minimum. Any spacing smaller than this d_0 will assure a Case I field for potential differences of absolute value greater than $|V_{c-t}|_{\min}$.

The limiting potential-distribution case was calculated for a molybdenum target by means of a simplified graphical analysis. For a primary-beam current density of 160 microamperes per square millimeter, the critical spacing for a primary electron energy of 375 volts and a grid bias of two volts was about one millimeter.

Experiment demonstrated that grid control of the secondary-electron current was satisfactorily linear for this spacing but was nonlinear for spacings three or more times greater.

The Case II field in the region $x_{\min}\leqslant x\leqslant d$ may be determined by further analysis quite similar to the foregoing. Remembering that, for this case, $\rho_-=0$, the mathematical procedure for this second region is the same as that present above, and yields the following formulas.

CASE II

$$x_{\min}\leqslant x\leqslant d,$$

$$\rho_{t2}=-\left(\frac{m}{2e}\right)^{\frac{1}{2}}\left\{\frac{J_{\text{pri}}}{(V+\phi)^{\frac{1}{2}}}+\int_{-\phi_{\min}}^{V_{\text{pri}}}\frac{J_\tau(eV)dV}{(V+\phi)^{\frac{1}{2}}}\right\}, \tag{15}$$

$$x=\int_{V_{c-t}}^{\phi}\frac{d\phi''}{+\left(\frac{2}{\epsilon}\int_{\phi_{\min}}^{\phi''}-\rho_{t2}(\phi')d\phi'\right)^{\frac{1}{2}}}+d. \tag{16}$$

Note that the integral will be negative and hence $x\leqslant d$. (The gradient, here represented by the square root, is positive in this region.) We then match the two Case II solutions for the separate regions at their common boundary $x=x_{\min}$, where $\phi=\phi_{\min}$.

$$\int_0^{\phi_{\min}}\frac{d\phi''}{-\left(\frac{2}{\epsilon}\int_{\phi_{\min}}^{\phi''}-\rho_t(\phi')d\phi'\right)^{\frac{1}{2}}}=\int_{V_{c-t}}^{\phi_{\min}}\frac{d\phi''}{+\left(\frac{2}{\epsilon}\int_{\phi_{\min}}^{\phi''}-\rho_{t2}(\phi')d\phi'\right)^{\frac{1}{2}}}+d, \tag{17}$$

where ρ_t is from Eq. (7) and ρ_{t2} is from Eq. (15). This equation is of the form $F_1(\phi_{\min})=F_2(\phi_{\min})$, and can be solved for the hitherto unknown constant $\phi_{\min}$. The $x_{\min}$ may be found by substitution of $\phi_{\min}$ back into either of Eqs. (13) or (16). ϕ may be found as a function of x by inverting Eqs. (12), (13), and (16) once they are obtained in algebraic form.

We thus have completed the determination of the potential distribution in the semi-infinite space-charge-filled region for both Case I and II potential distributions. To summarize, the solution for a Case I field is given by use of Eqs. (7), (12), and (14), and the solution of Case II is given by Eqs. (7), (13), (15), and (17)—all in conjunction with Eqs. (1) through (5).

If desired, the net current density at any position in the interelectrode region may be easily calculated. Due

to the continuity of current, the net current density J_n is the same at any point in the semi-infinite region, and is given by Eq. (18).

$$J_n=\int_{-\phi_{\min}}^{V_{pri}} J_\tau(eV)dV-J_p,$$
$$J_n=J_p(\delta-1)-\int_0^{-\phi_{\min}} J_\tau(eV)dV. \tag{18}$$

It is interesting to note that J_n may be zero, although the interelectrode region is filled with moving electrons. This consideration played an important role in the electron tube for which these calculations were originally made.

Finally, the methods and results of this analysis are not restricted to treating only secondary electrons but may be extended to include other types of electron emission (assuming that the angular distribution is a cosine function) if the corresponding energy distributions are known. It is then only necessary to substitute the new distribution functions and alter the appropriate limits of integration.

GLOSSARY OF SYMBOLS

d = separation of grid and target (meters).
$d\sigma$ = differential area emitting secondary electrons.
δ = secondary-electron yield ratio.
e = charge of electrons, 1.602×10^{-19} coulomb.
ϵ = dielectric constant of free space, 8.854×10^{-12} coulomb/meter.
$i(eV)$ = secondary-electron energy distribution.
J = current density.
$J(eV)$ = current-density secondary-electron energy distribution.
$J(v)$ = current-density secondary-electron speed distribution.
$J(v_x)$ = current-density secondary-electron normal velocity component distribution.
$J_\tau(eV)$ = current-density energy distribution of the equivalent secondary-electron population [see Eq. (5)].
J_{pri} = primary-electron current density.
m = mass of electrons, 9.11×10^{-31} kg.
ϕ = electric potential.
ϕ' and ϕ'' = variables of integration.
$\phi_{\min}$ = minimum electric potential between electrodes.
q = substitution variable $=d\phi/dx$.
ρ = charge density.
ρ_{pri} = primary-electron charge density.
ρ_+ = outgoing secondary-electron charge density.
ρ_- = reflected secondary-electron charge density.
ρ_τ = total charge density.
θ = angle of electron emergence, measured from electrodes' normal.
V = voltage equivalent of electron energy (eV).
V_{pri} = voltage equivalent of primary-electron energy.
v = electron speed.
x = distance measured along normal from target to grid.

Methods of Processing Silver-Magnesium Secondary Emitters for Electron Tubes

PAUL RAPPAPORT
RCA Laboratories Division, Princeton, New Jersey
(Received June 8, 1953)

In order to prepare a good secondary electron emitting surface from a silver-magnesium alloy, it was found that baking in dry oxygen would not form the desired surface as originally believed, whereas baking in the presence of water vapor did yield a good secondary emitter. Subsequent experiments showed that, because of the rapid diffusion of oxygen through heated silver, the magnesium in the alloy was oxidized in situ without diffusing to the surface to form the required magnesium oxide layer. With water vapor as an oxidizing medium, the diffusion rate of magnesium through the silver greatly exceeded that of the water vapor, and a satisfactory surface layer was formed. A new technique was then devised in which an original water-vapor process formed the desired surface, followed by an oxygen process to oxidize the remanent magnesium throughout the volume of the alloy. With the new technique, good secondary emitters are possible which will withstand severe overheating without the evaporation of magnesium found with earlier techniques.

I. INTRODUCTION

IN general, the best-known secondary electron emitters are those that consist of an oxide film on a metal electrode base. Such an emitter is obtained when an alloy of 98.3 percent silver and 1.7 percent magnesium[1] is baked for a suitable time in an oxidizing atmosphere. Figure 1 shows a drawing indicating schematically the structure associated with an activated emitter of this kind. The surface film has been identified as magnesium oxide by electron-diffraction studies and its thickness measured to be about 1250 angstrom units.[2] Properly prepared silver-magnesium (AgMg) alloy surfaces have the following excellent characteristics for use in vacuum tubes:

(1) High secondary-emission ratios that remain constant for several thousand hours.
(2) Low surface and interface resistances, so that arcing and Malter effect[3] do not occur, thus giving good high-frequency performance.
(3) Ease and uniformity of activation of the alloy once the proper parameters are set.
(4) No deterioration of the activated sample upon a few hours' exposure to air.
(5) No further activation required after the emitter is placed in the vacuum tube.

Two difficulties connected with this type of emitter are its inability to withstand heat because of evaporation of the magnesium remaining in the alloy, and the poor secondary-emission ratios that many workers experience because of incorrect application of the processing technique.

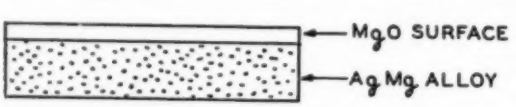


FIG. 1. Silver-magnesium secondary-emitting surface showing thin MgO film.

The purpose of the present work was to understand the mechanism of the oxide-film formation on the silver-magnesium alloy. It is believed that, at elevated temperatures, the magnesium becomes mobile in the silver and some of it diffuses to the surface of the alloy. There it will either evaporate or combine chemically and thus create a diffusion gradient to the surface for magnesium ions. Since magnesium oxide is the desired product on the surface of the alloy, it was originally expected that, if the heated alloy were exposed to oxygen, a substantial oxide film would form. However, this was generally not the case. On the other hand, use of other oxidizing media, such as water vapor, gave much better oxide films, and it was recognized that a precise understanding of the processing mechanism was lacking.

II. EXPERIMENTAL TECHNIQUES

An empirical approach to the problem was undertaken. Processing experiments were performed by baking the five-mil-thick alloy in various atmospheres for various times and temperatures, with separate control of the important parameters. The atmospheres used included oxygen, air, nitrogen, helium, hydrogen, and water vapor. The pressure was varied from atmospheric to 10^{-6} mm/Hg. Time of heating varied from a few minutes to four hours and the temperature from about 400 to 700°C.

Great care was taken in the experiments to eliminate equivocal results wherever possible. Baking was almost always done in a quartz tube which was evacuated by an ordinary mechanical pump and a mercury diffusion pump. When necessary, the gases introduced into the processing system were dried through cascaded liquid air traps and, for purity of certain gases, clean charcoal traps at liquid air temperatures were employed. It was sometimes necessary to flush out the system a number of times to insure purity of the baking atmosphere. Special polishing and cleaning techniques for the alloy

[1] Zworykin, Ruedy, and Pike, J. Appl. Phys. **12**, 696–698 (1941).
[2] P. . Rappaport, Massachusetts Institute of Technology Eleventh Annual Conference on Physical Electronics Report, pp. 70–73 (1951).
[3] L. Malter, Phys. Rev. **50**, 48–58 (1936).

were employed, as follows. After having been polished on a standard clean cloth buffing wheel, the sample was washed in a soapy water solution, rinsed in distilled water, followed by a rinse in distilled acetone and methanol mixture.

The results of the various processes were evaluated by the physical and electrical properties of the treated samples. Appearance, mechanical flexibility, electron and x-ray diffraction, evaporation of heated samples, secondary-emission ratio, and life tests all gave important clues. A test was developed in which a drop of mercury was applied to the surface of a processed sample and then withdrawn. By noting the difference in wetting action, an oxidized surface can be differentiated from an unoxidized one. For example, if an unprocessed sample of silver-magnesium alloy is contacted by the mercury drop for even a few seconds, some of the mercury will stick to it, while an oxidized sample remains free from any mercury even though there is contact for several minutes.

III. EXPERIMENTAL RESULTS

Table I lists the important characteristics of the AgMg samples after they were baked in the various atmospheres. The treatments of the alloy are listed in the vertical column of the table. The standard processing technique, referred to as process No. 1, is the most reliable and reproducible method for the processing of AgMg into good secondary emitters. It was developed by C. W. Mueller at the RCA Laboratories and consists essentially of baking the clean alloy at 550°C for ½ hour in a quartz oven at about 10^{-5} mm/Hg using a dry-ice and acetone cold trap. Before sealing, water vapor is introduced into the system by bubbling air through water into the trap, previous to applying the dry ice and acetone. The characteristics of such a sample, listed in the table, are all good except for the fact that magnesium evaporates when the sample is heated.

The treatments from process No. 1 to No. 6 were all in the same chamber and with baking for ½ hour at 550°C. Water vapor was added intentionally only in process No. 1. In process No. 2, a liquid-air trap was used and the system pumped down to about 10^{-6} mm/Hg before baking. No detectable change of the alloy took place except for a general out-gassing and cleaning. In process No. 3, the sample was baked in air with no drying, and in process No. 4 the system was pumped down to about 10 microns before baking took place. Process No. 5 took place in pure dry oxygen at atmospheric pressure. In process No. 6 the chamber was filled with helium directly from a tank with no drying precautions, whereas in process No. 7 charcoal-in-liquid-air traps were used, and the system was flushed out three times before baking. Nitrogen directly from a tank was used in process No. 8, and in No. 9 a sample was baked in a so-called "dry" hydrogen furnace at atmospheric pressure for 5 to 10 minutes at about 600°C.

If a sample is processed as in No. 10 of the table, which is process No. 1 followed by process No. 5, it becomes brittle but has a good ratio as shown in the table. This so-called two-step process will be discussed later in another connection.

The top row of the table lists the characteristics of the processed alloy. The photograph in Fig. 2 shows the typical appearance of processed samples. The upper one shows what is meant by patchy, and the lower one is the shiny yellow referred to in the table. These are microphotographs magnified about 200 times. Polish lines are visible on both samples. The shiny yellow oxide film on the lower photograph is transparent so that the underlying alloy is visible. The white spots on the upper photograph are probably dots of MgO appearing to the unaided eye as a milky white film on

TABLE I. Characteristics of AgMg alloy (1.7 percent Mg) samples when processed by various methods.

Process no.	Baking atmosphere[a]	Water vapor pressure mm/Hg	Secondary-emission ratio	Appearance	Presence of oxide film (mercury test)	Flexibility	Magnesium evaporation	Further oxidation possibilities
1	Vacuum (standard process)	$\approx 10^{-4}$	Good	Shiny yellow film	Yes	Yes	Much	Yes
2	Vacuum (liq. air trap)	$<10^{-6}$	Poor	Metallic	No	Yes	Much	Yes
3	Air 760 mm/Hg	No drying	Poor	Patchy dull white film	No	(Very brittle) No	Very little	No
4	Air 10^{-2} mm/Hg	No drying	Poor	Patchy dull white film	No	(Less brittle) No	Little	No
5	Oxygen 760 mm/Hg	$<10^{-6}$	Poor	Patchy dull white film	No	(Very brittle) No	Very little	No
6	Helium wet	No drying	Good	Shiny yellow film	Yes	Yes	Much	Yes
7	Helium dry	$<10^{-5}$	Poor	Metallic	No	Yes	Much	Yes
8	Nitrogen	No drying	Not tested	Patchy dull white film	No	?	?	?
9	"Dry" hydrogen	8×10^{-3}	Good	Shiny yellow film	Yes	Yes	Much	Yes
10	Two-step process	$\approx 10^{-4}$	Good	Shiny yellow film	Yes	(Brittle) No	Very little	No

[a] Baking done at 550°C for ½ hour for all processes except 9 and 10.

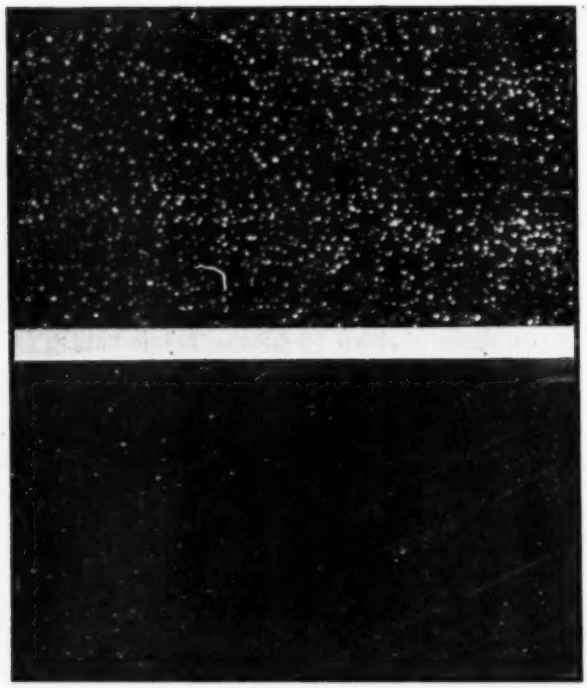


FIG. 2. Microphotographs (magnified 200 times) of processed AgMg surfaces. Top—sample baked in oxygen as in process No. 3, 4, and 5. Below—sample processed in water vapor as in process No. 1.

the surface of the alloy. The dark regions are probably silver, as indicated by the mercury drop test. The drop immediately wets the surface which must mean that some silver base metal is exposed at the surface. Tests with H_2S verified these results. The silver regions blackened because of the formation of silver sulfide. Comparisons with crystal-grain boundaries did not show any correlation with the white dots. The dots are much more numerous and more evenly spaced than the grain boundaries.

The test for flexibility is important because it gives information concerning the internal structure and composition of the sample. In the table, "yes" means that the flexibility is the same as that of the original AgMg alloy. When the samples became brittle, they were incapable of any flexing, breaking much the same as a razor blade would.

A secondary emission ratio of 3.5 or over is considered good in the table. A ratio of 1.2 is the approximate value observed for unprocessed AgMg alloy. Readings in the range from 1 to 2 are considered poor for a processed surface. Tests of secondary emission were made in rotating dynode tubes[4] where a number of surfaces are all tested in the same bulb so as to eliminate variations due to different tube processing.

Tests for magnesium evaporation were conducted on samples placed in vacuum bulbs and then heated by radio frequency to about 775°C. The amount of evaporation was estimated qualitatively by the light transmission of the metallic film appearing on the inside bulb wall. Figure 3 shows two bulbs tested in this manner. In the table "very much" magnesium evaporation refers to deposits that were from about 75 percent transmitting to almost opaque, while "very little" refers to deposits that were barely visible, affecting the transmission of light no more than a few percent.

The ease with which a sample may be reprocessed, (i.e., further formation of an oxide film on a sample that has been previously processed) indicates something about the physical and chemical state of the processed sample. This will be made clearer in the discussion. According to the table, after process No. 1 a sample can be made darker by additional oxidation. However, this can be done only within narrow limits. Surface-film thickness measurements indicate a variation from 1150A to 1350A for extremes in time and temperature of processing. Of course, after process No. 2, vacuum baking, the sample can be oxidized. The interesting thing is that after processes No. 3, 4, 5, 8, and 10, no further oxidation is possible.

Note that the characteristics of samples treated as in processes No. 1, 6, and 9 are very similar. A processed sample can be made almost identical by any one of the three processes if the conditions such as temperature and time of firing, and water-vapor content, are properly controlled.

Some further important experiments were performed, as follows:

If a sample of AgMg alloy is coated by evaporation with a thin gold film about 3000A thick and then baked in oxygen, as in process No. 5, it remains flexible, and a magnesium-oxide layer is produced on the surface. This is remarkably different from the nongold coated sample of AgMg. Further processing was possible by process 1. The sample was not tested for secondary emission.

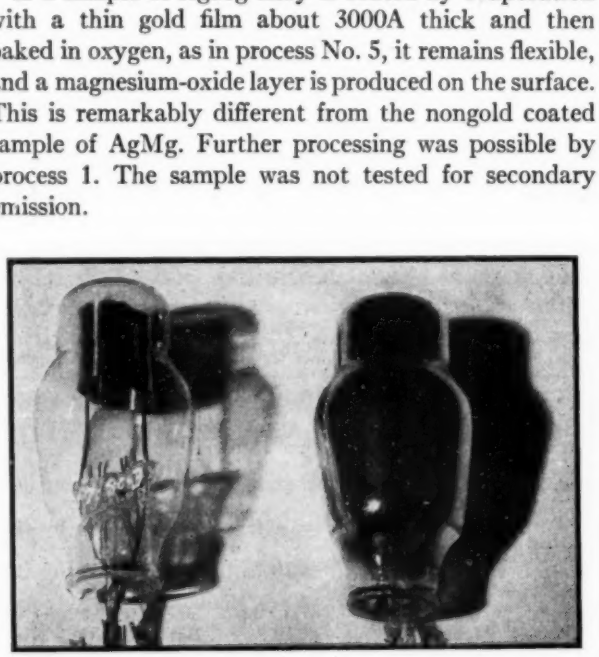


FIG. 3. Appearance of vacuum bulb after samples contained have been "rf" heated to 775°C. Tube on right contains a normally processed AgMg emitter heated at 775°C for 15 sec. Tube on left contains a two-step processed emitter heated at 775°C for four min.

[4] C. W. Mueller, Proc. Inst. Radio Engrs. 38, 159–164 (1950).

If a sample of AgMg alloy is coated with silver about 1000A thick, it is still possible to form an oxide on its surface by process No. 1, 6, or 9. Further, if the once-processed oxidized sample is again coated with a silver film and processed again by process No. 1, a second oxide coating is formed which has the characteristics of line 1 in the table, including a good secondary-emission ratio.

IV. DISCUSSION OF RESULTS

The striking conclusions that can be drawn from these experiments are: When samples are baked in oxygen-containing atmospheres as in processes No. 3, 4, and 5, very poor secondary emitters result, but when baking in atmospheres where oxygen is not present and water vapor is, as in treatments No. 1, 6, and 9, good emitters are obtained. This is opposite to the expectation that oxygen should form the MgO film and give good secondary emitters.

The following qualitative picture is supported by the experimental results. It is well known that oxygen diffuses through hot silver in large quantities.[5,6] The addition of small amounts of magnesium does not appear to affect this phenomenon very much. Magnesium diffuses through the hot silver readily but not nearly so fast as the oxygen. Magnesium-oxide and water-vapor molecules do not diffuse appreciably through silver because of their large size. The present experiments have shown that oxygen and magnesium also diffuse through the thin MgO surface when formed on the alloy. When the AgMg alloy is baked in an oxygen atmosphere, the oxygen diffuses into the alloy much faster than the Mg can diffuse to the surface. As the oxygen penetrates the alloy, it combines with any magnesium atoms that it contacts. The MgO formed is immobile and is frozen in place, thus preventing any formation of a surface MgO film, except where Mg atoms are on or near the surface, about 1 out of every 13 atoms.

The brittleness of the alloy prepared with oxygen results from the effect of the internally formed MgO on the silver lattice. The patchy appearance of the surface, the mercury drop test, and the secondary-emission tests all confirm that a continuous MgO film does not form on the surface. Finally, since it is not possible to oxidize the sample further or evaporate any magnesium from it, it is concluded that substantially all of the magnesium in the sample has been oxidized without producing a good secondary-emitting surface layer.

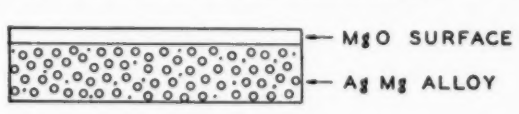


FIG. 4. Silver-magnesium secondary-emitting surface after the two-step processing. The circles represent MgO molecules and the dots the remaining magnesium atoms.

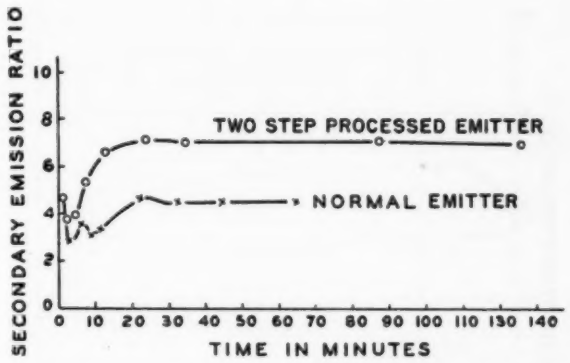


FIG. 5. Secondary emission at 300 v showing effect of two-step processing of AgMg alloy. Emitters in same bulb on demountable system.

When the alloy is baked in an oxygen-free atmosphere with some other oxidizer such as water vapor, which does not readily diffuse into the alloy, the picture is different. In this case, the magnesium is allowed to diffuse to the surface before it is oxidized by the action of water vapor and heat, and a good secondary-emitting layer is formed. That free magnesium is present inside such a sample is evident from the high evaporation rate when a water-vapor processed sample is heated and also from its ability to form further oxides on top of evaporated silver films. Water-vapor processed samples are mechanically as flexible as the original AgMg alloy. The results in Table I demonstrate the effect of water vapor in the processing atmosphere. The interesting point about the "dry" hydrogen process is that free oxygen cannot exist and can definitely be eliminated as a contributor to the oxide-film formation.

The results with the gold-covered AgMg sample which did not become brittle and was capable of reprocessing after baking in oxygen can now be explained. Since oxygen diffuses through gold slightly, compared to silver, none of the internal free magnesium was oxidized. The magnesium now diffuses through the gold film whenever it is heated and becomes oxidized on the surface.

V. REDUCTION OF MAGNESIUM EVAPORATION

Since the water vapor when used as the oxidizing agent cannot penetrate the alloy, any magnesium that has not arrived at the surface is still present in metallic form, inside the volume of the alloy. This magnesium can evaporate if the emitter is overheated, as is readily possible during processing or operation of electron tubes. This is highly undesirable because the evaporation of magnesium may cause leakage and rf losses.

An extension of the above techniques can cure this undesirable property. The experiments here described have shown that, when a water-vapor processed AgMg sample is subsequently heated in oxygen, it becomes brittle, and evaporation of magnesium after over-

[5] E. W. R. Steacie and F. M. Johnson, Proc. Roy. Soc. (London) A112, 542 (1926).
[6] F. M. Johnson and P. Larose, J. Am. Chem. Soc. 46, 1377–1389 (1924).

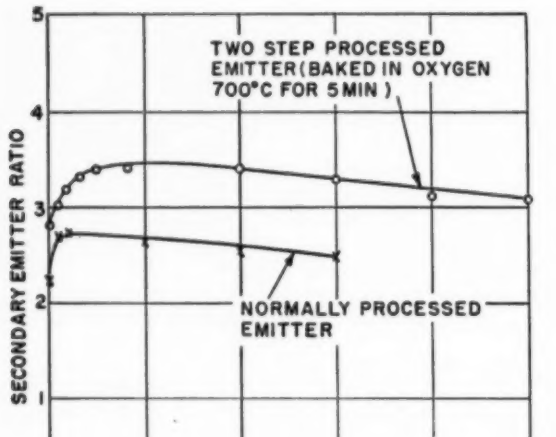


FIG. 6. Comparison of life tests of AgMg secondary emitters in same sealed-off and gettered bulb.

heating is almost completely eliminated. The effect indicates that oxygen penetrates the already formed MgO film and oxidizes much of the remaining free magnesium atoms in the alloy, thus freezing them in place. Figure 4 shows a simplified illustration of such an emitter after the two-step activation process.

Measurements, in a vacuum of about 10^{-5} mm/Hg, were made to determine the amount of magnesium evaporation at 775°C from the two-step activated sample by comparison with evaporation from samples of AgMg with known Mg concentration. The results show that when a water-vapor activated sample is subsequently baked at 700°C in oxygen at atmospheric pressure for 5 min, the magnesium evaporation in a vacuum is about one-twentieth that of a similar sample without oxygen bake. Figure 3 shows two bulbs containing geometrically similar samples that were heated to 775°C. The bulb on the right contains a normally processed sample (water vapor followed by oxygen). Even in the photograph, the difference in evaporation is striking. Conditions could probably be set to freeze an even greater amount of the magnesium atoms in place by this method. Figure 5 shows the secondary-emission ratio of the two-step activated sample in comparison with the one-step activation both in the same demountable bulb. A significant increase in ratio is observed by this two-step treatment, probably because of the effect of the oxygen in the MgO surface. Some free Mg in the surface may be oxidized which would increase the mean free path for secondary electrons and give rise to a greater yield. Figure 6 shows comparisons of 200-hr life tests in the same sealed-off bulb. It should be noted that the ratios in Fig. 6 are lower than those in Fig. 5. This is often observed when comparing results between demountable and sealed-off tubes. The surfaces are apparently sensitive to small variations in tube atmosphere.

VI. CONCLUSIONS

In conclusion, a description has been presented of what takes place when silver magnesium alloy is processed for making secondary emitters. In general, oxygen has been shown to be detrimental to the initial purpose of building an MgO film on the alloy because of its solubility in hot silver. Water vapor, because of its oxidizing ability and its inability to diffuse into the alloy, has given successful, reproducible, long-life, and high-ratio secondary emitters. Further application of the principles has resulted in a two-step activation process which gives an even higher ratio secondary emitter whose tolerance, with respect to overheating, is far superior to existing AgMg secondary-emitting surfaces.

VII. ACKNOWLEDGMENTS

The writer wishes to express his appreciation to Dr. C. W. Mueller of the RCA Laboratories for his helpful contributions during the course of this work and to Mrs. E. G. Moonan for her careful processing of the many samples used in these experiments.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 3 MARCH, 1954

Symmetry as a Factor in Finite Difference Approximations

C. M. FOWLER
Physics Department, Kansas State College, Manhattan, Kansas
(Received June 8, 1953; revised manuscript received September 30, 1953)

The effect of symmetry of the solutions to the finite difference diffusion equation upon the stability criterion is discussed. A simple example is included to illustrate the discussion.

IT is well known that finite difference approximations of diffusion equation solutions will oscillate and ultimately diverge in time, unless certain precautions are observed with regard to the sizes of time and space increments employed. This feature is of importance in machine calculations. The present note discusses one factor, not normally considered explicitly, that affects the convergence of the approximate solutions, namely, the symmetry of the solutions.

Consider the difference equation applicable to the one-dimensional diffusion equation with constant physical parameters.[1]

$$T_{x,t}=[T_{x+1,t-1}+(M-2)T_{x,t-1}+T_{x-1,t-1}]/M, \quad (1)$$

$$\Delta t=(\Delta x)^2/M\alpha. \quad (2)$$

In these equations, Δx and Δt are the finite space and time increments selected for the problem, while the integers x and t serve as independent variables and denote, respectively, the position of a point in units of Δx and the time in units of Δt. The dependent variable $T_{x,t}$ may represent temperature or concentration at (x,t), for example. The quantity α is the appropriate diffusivity.

Dusinberre[1] appears to have been the first to introduce the general relationship of Eq. (2) which relates the time and space increments by an arbitrary number, M, referred to as the modulus. It would appear from this equation that quite arbitrary values of Δt could be selected for a given choice of space increment, by suitably choosing M. This would be a very valuable feature, since the choice of a small M would result in a large value of Δt. Thus, the number of arithmetic operations needed to carry a solution to a predetermined value of time would be reduced. It is found, however, that if M is chosen too small, the solutions ultimately diverge.

The question of minimum modulus has been the subject of several discussions.[1,2,3] It is known that the minimum modulus is influenced by the particular boundary conditions as well as by the number of spacings l selected for the problem. Consideration of the following simple example shows that the symmetry of the solutions is also a factor.

Suppose the slab, of thickness l, has the symmetric boundary conditions $T_{0,t}=T_{l,t}=0$, while the initial conditions, also symmetric, are $T_{x,0}=F(x)=F(l-x)$. It is not difficult to show that the solution of this problem is given by[2]

$$T_{x,t}=\sum_{n=0}^{l} A_n\left(1-\frac{4}{M}\sin^2\frac{n\pi}{2l}\right)^t \sin\frac{n\pi x}{l}, \quad (3)$$

$$A_n=\frac{2}{l}\sum_{x=0}^{l} F(x)\sin\frac{n\pi x}{l}. \quad (4)$$

For time convergence, we must have

$$\left|1-\frac{4}{M}\sin^2\frac{n\pi}{2l}\right|\leq 1.$$

This requires that

$$M\geqslant 2\sin^2\frac{n\pi}{2l} \quad (5)$$

(for all n required in the solution).

For the symmetric solution cited, it is clear from Eq. (4) that all A_n are zero for even n. Equation (5) will then be satisfied generally if it holds for the odd integer nearest to l. These integers are $l-1$ if l is even, and l if l is odd.

The conditions for convergence then become

$$\begin{aligned} &M\geqslant 2\cos^2\pi/2l; && l \text{ even}\\ &M\geqslant 2; && l \text{ odd}. \end{aligned} \quad (6)$$

Thus, the modulus may be lowered if the solution is symmetric. It is clear that had the solution not had the symmetry, both even and odd A_n would have entered the solution, and the larger value of the modulus would have been required for convergence.

The above analysis applies to solutions possessing even parity about the slab midpoint. A corresponding analysis can be made for solutions of odd parity, and for other higher-symmetry types, although such solutions are not of common occurrence in practical diffusion problems.

A similar analysis shows the same general behavior to hold for other types of boundary conditions, such as those treated in reference 2. The eigenfunctions and

[1] G. M. Dusinberre, Trans. Am. Soc. Mech. Engrs. 67, 703–709 (1945).
[2] C. M. Fowler, Quart. Appl. Math., Vol. III, No. 4, 361–376 (1946).
[3] W. B. Van Arsdel, Report No. ED-6-12-19, Western Regional Research Laboratory (1948).

eigenvalues required for the series expansions are more complicated in general, but particular symmetries will allow lower moduli. Actually, the boundary conditions given for the simple example above invert the role of even and odd l because of the accidental vanishing of the space portion of the eigenfunction when $n=l$. This, of course, is not the general situation.

By a straightforward extension of the above arguments, the symmetry effect can be shown to persist for two- and three-dimensional rectangular bodies.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 3 MARCH, 1954

Approximations in Linear Viscoelasticity Theory: Delta Function Approximations*

HERBERT LEADERMAN
National Bureau of Standards, Washington, D. C.
(Received June 12, 1953)

The relationships between certain approximations that have been proposed recently in connection with the linear viscoelastic behavior of noncrosslinked amorphous polymers are discussed. The approximations reviewed here are applicable with slight modifications to other types of linear response, for example, dielectric behavior.

THE response of a material manifesting linear viscoelastic behavior to a simple shear stress varying with time in an arbitrary manner is given, assuming the Superposition Principle, by an integral equation of the convolution type. The shear strain is given by

$$\gamma(t)=\int_{-\infty}^{0}\frac{d\sigma(t-k)}{dk}J(k)dk. \tag{1}$$

For a noncrosslinked amorphous polymer,

$$J(t)=J_0+J\psi(t)+t/\eta \tag{1a}$$

is the creep compliance at time t due to a unit shear stress applied at zero time; σ is the shear stress and J_0, J, and η are constants of the material (at constant temperature). Since $\psi(t)$ is a nondecreasing function, it can be represented in the following form:

$$1-\psi(t)=\int_{-\infty}^{\infty}e^{-t/\tau}L(\ln\tau/\tau_0)d(\ln\tau/\tau_0), \tag{2}$$

where

$$\int_{-\infty}^{\infty}L(\ln\tau/\tau_0)d(\ln\tau/\tau_0)=1.$$

Then the dynamic response is given by

$$\{J'(\omega)-J_0\}/J=\alpha(\omega)=\int_{-\infty}^{\infty}\frac{L(\ln\tau/\tau_0)d(\ln\tau/\tau_0)}{1+\omega^2\tau^2}, \tag{3}$$

$$\{J''(\omega)-1/\omega\eta\}/J=\beta(\omega)=\int_{-\infty}^{\infty}\frac{\omega\tau L(\ln\tau/\tau_0)d(\ln\tau/\tau_0)}{1+\omega^2\tau^2}. \tag{4}$$

The functions $\psi(t)$, $\alpha(\omega)$, and $\beta(\omega)$ are computed directly from $L(\ln\tau/\tau_0)$ by means of the above equations.

Certain approximations are necessary to obtain $L(\ln\tau/\tau_0)$ from experimental data available in graphical or numerical form. One type of approximation involves the use of either delta-functions or cut-off integrals. These methods will be shown to be essentially equivalent. The first use of such methods appears to be due to Gevers and du Pré[1] in connection with the analogous dielectric problem. In the polymer field, Alfrey and Doty[2] have given the results of the application of this technique to Eq. (2). This approximation has been discussed by Ferry,[3] Andrews,[4] and Schwarzl.[5–8]

The earlier presentation of Gevers and du Pré seems more rigorous than certain of the later ones. For this reason, it seems to be of value to present these approximations by the methods of Gevers and du Pré.

With the notation

$$z=\ln\tau/\tau_0,$$
$$n=\ln t/\tau_0, \quad \text{or} \quad -\ln\omega\tau_0,$$

Eqs. (2), (3), and (4) become, respectively,

$$\psi(n)=\int_{-\infty}^{\infty}[1-\exp(-e^{n-z})]L(z)dz, \tag{5}$$

$$\alpha(n)=(\tfrac{1}{2})\int_{-\infty}^{\infty}[1+\tanh(n-z)]L(z)dz, \tag{6}$$

$$\beta(n)=(\tfrac{1}{2})\int_{-\infty}^{\infty}\operatorname{sech}(n-z)L(z)dz. \tag{7}$$

* This paper presents the results of work performed as part of a project supported by the U. S. Office of Naval Research and Office of the Quartermaster General.

[1] M. Gevers and F. K. du Pré, Trans. Faraday Soc. **A42**, 47 (1946).
[2] T. Alfrey and P. Doty, J. Appl. Phys., **16**, 700 (1945).
[3] Ferry, Sawyer, Browning, and Groth, J. Appl. Phys. **21**, 513 (1950).
[4] R. D. Andrews, Ind. Eng. Chem. **44**, 707 (1952).
[5] F. Schwarzl, Physica **17**, 830 (1951).
[6] F. Schwarzl, Physica **17**, 923 (1951).
[7] F. Schwarzl and A. J. Staverman, Physica **18**, 791 (1952).
[8] F. Schwarzl, paper presented at the Second International Congress on Rheology, Oxford, England, July, 1953. The author is indebted to Dr. Schwarzl for an advance copy of this paper.

DELTA FUNCTION APPROXIMATIONS

Differentiating Eq. (5) with respect to n, we obtain

$$d\psi(n)/dn=\int_{-\infty}^{\infty}\exp(n-z)\exp(-e^{n-z})L(z)dz. \quad (8)$$

The function

$$\exp(n-z)\exp(-e^{n-z})$$

can be considered as an approximation to a delta function. It possesses a maximum at $z=n$, and falls off rapidly on either side of this maximum. Thus it differs significantly from zero only over a small range of values of z around $z=n$. Over this range $L(z)$ may be considered constant to a first approximation and equal to its value at $z=n$. This is denoted by $L(n)$. Thus Eq. (8) becomes

$$d\psi(n)/dn\simeq L(n)\int_{-\infty}^{\infty}\exp(n-z)\exp(-e^{n-z})dz. \quad (9)$$

The value of the integral is unity, hence

$$d\psi(n)/dn\simeq L(n) \quad (10)$$

which is Alfrey's approximation.[2]

Consider now the function

$$f(n)=4\int_{-\infty}^{\infty}[\exp(n-z)\exp(-e^{n-z})]^2L(z)dz. \quad (11)$$

Since the function

$$[\exp(n-z)\exp(-e^{n-z})]^2$$

is a closer approximation to a delta function than the one previously considered, it is clear that, as before, $L(z)$ may be replaced by $L(n)$ outside the integral. We then have the approximate relationship

$$\left.\begin{aligned} f(n)&\simeq 4L(n)\int_{-\infty}^{\infty}[\exp(n-z)\exp(-e^{n-z})]^2dz \\ &\simeq L(n). \end{aligned}\right\} \quad (12)$$

This is a better approximation than the Alfrey approximation, since the second function of $(n-z)$ is a better approximation to a delta function. Now Eq. (11) may be written

$$f(n)=\int_0^{\infty}(2t/\tau)^2\exp(-2t/\tau)L(\ln\tau/\tau_0)(d\tau/\tau), \quad (11a)$$

and from Eq. (5) we obtain

$$1-\psi(2t/\tau_0)=\int_0^{\infty}\exp(-2t/\tau)L(\ln\tau/\tau_0)(d\tau/\tau). \quad (5a)$$

It is seen that

$$\left(\frac{2t}{\tau_0}\right)^2\frac{d^2\psi(2t/\tau_0)}{d(2t/\tau_0)^2}=-f(n),$$

from which is obtained the relation given by Schwarzl:[7]

$$f(n)=d\psi(m)/dm-d^2\psi(m)/dm^2, \quad (13)$$

where

$$m=n+\ln 2=\ln(2t/\tau_0).$$

In general, Schwarzl's second approximation, given by Eq. (13) would be expected to be a closer approximation than Alfrey's approximation. In a similar way the third approximation obtained from

$$g(n)=(27/2)\int_{-\infty}^{\infty}[\exp(n-z)\exp(-e^{n-z})]^3L(z)dz \quad (14)$$

is better than the second approximation.

Now consider Eqs. (6) and (7). Differentiating the former with respect to n, we obtain

$$d\alpha(n)/dn=(\tfrac{1}{2})\int_{-\infty}^{\infty}\operatorname{sech}^2(n-z)L(z)dz. \quad (15)$$

Here again, $\operatorname{sech}^2(n-z)$ is an approximation to a delta function, and differs from zero only over a small range of values around $n=z$. Thus we may write approximately:

$$\left.\begin{aligned} d\alpha(n)/dn&\simeq(\tfrac{1}{2})L(n)\int_{-\infty}^{\infty}\operatorname{sech}^2(n-z)dz \\ &\simeq L(n) \end{aligned}\right\}. \quad (16)$$

Similarly, $\operatorname{sech}(n-z)$ may be considered very approximately as a delta function. We thus obtain from Eq. (7), the crude approximation[8]

$$\beta(n)\simeq(\tfrac{1}{2})L(n)\int_{-\infty}^{\infty}\operatorname{sech}(n-z)dz. \quad (17)$$

Hence,

$$L(n)\simeq(2/\pi)\beta(n).$$

Equations (17) and (16) may be considered as giving, respectively, first and second approximations to $L(n)$. It is clear that the functions $p(n)$ and $q(n)$ given by

$$p(n)=(2/\pi)\int_{-\infty}^{\infty}\operatorname{sech}^3(n-z)L(z)dz \quad (18)$$

and

$$q(n)=(4/3)\int_{-\infty}^{\infty}\operatorname{sech}^4(n-z)L(z)dz \quad (19)$$

can be used to obtain third and fourth approximations to $L(n)$, provided that $p(n)$ and $q(n)$ can be expressed in terms of measurable quantities. These turn out to be the higher approximations given by Schwarzl:[8]

$$p(n)=(2/\pi)[\beta(n)-d^2\beta(n)/dn^2], \tag{20}$$

$$q(n)=d\alpha(n)/dn-(\tfrac{1}{4})d^3\alpha(n)/dn^3. \tag{21}$$

CUT-OFF INTEGRAL APPROXIMATIONS

Considering Eq. (5), it is seen that

$$1-\exp(-e^{n-z})\simeq 1 \quad \text{when} \quad z<n$$
$$\simeq 0 \quad \text{when} \quad z>n.$$

The function changes in value from zero to unity over a narrow range of values of z around $z=n$. The above function may be considered approximately as a step function. Hence,

$$\psi(n)\simeq\int_{-\infty}^{n} L(z)dz. \tag{22}$$

Differentiating Eq. (22) with respect to n we obtain

$$d\psi(n)/dn\simeq L(n)$$

as before. A similar procedure may be applied to $\alpha(n)$ and $\int\beta(n)dn$. Since the integral of the delta function is the step function, the integral of an approximation to the delta function is an approximation to the step function, hence the two approximation methods are equivalent.

STRESS RELAXATION, STORAGE MODULUS, AND LOSS MODULUS

Similar approximation methods may be applied to data obtained from stress relaxation experiments and from dynamic data expressed as storage modulus and loss modulus. Corresponding to Eq. (1), the shear stress as a function of shear strain history may be written

$$\sigma(t)=\int_{-\infty}^{0}\frac{d\gamma(t-k)}{dk}G(k)dk. \tag{23}$$

In a stress relaxation experiment on an amorphous polymer, in which unit shear strain is applied at zero time, the shear stress at time t is $G(t)$ where

$$G(t)=G+G_0\phi(t) \quad \text{for crosslinked polymers,}$$
$$=G_0\phi(t) \quad \text{for noncrosslinked polymers.} \tag{24}$$

Let

$$\phi(n)=\int_{-\infty}^{\infty}\exp(-e^{n-z})M(z)dz, \tag{25}$$

where

$$\int_{-\infty}^{\infty}M(z)dz=1.$$

Then by differentiating Eq. (25) with respect to n we obtain

$$-d\phi(n)/dn\simeq M(n). \tag{26}$$

Let the storage modulus and loss modulus be given by $G'(\omega)$ and $G''(\omega)$, respectively. Then

$$\{G'(n)-G\}/G_0=\lambda(n)$$
$$=(\tfrac{1}{2})\int_{-\infty}^{\infty}[1-\tanh(n-z)]M(z)dz \tag{27}$$

and

$$G''(n)/G_0=\mu(n)=(\tfrac{1}{2})\int_{-\infty}^{\infty}\operatorname{sech}(n-z)M(z)dz. \tag{28}$$

Hence we obtain, as before,

$$M(n)\simeq -d\lambda(n)/dn \tag{29}$$

$$M(n)\simeq(2/\pi)\mu(n) \tag{30}$$

and the same higher approximation techniques are still valid.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 3 MARCH, 1954

RLC Canonic Forms*†

F. M. REZA
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received June 12, 1953)

Foster's reactance theorem synthesizes the class of lossless networks in a so-called canonic form. This idea is here generalized in order to show that there is a distinct class of RLC networks possessing a canonic form. It will be shown that when an RLC driving-point impedance or admittance possesses a canonic form, its poles and zeros must alternate on each and every "separate part" of an algebraic curve (c) of a special nature located in the left half of the frequency plane $s=\sigma+j\omega$, belonging to one of the following two families:

$$a_0+\sum_1^n \frac{a_k(\sigma+\alpha_k)}{(\sigma+\alpha_k)^2+\omega^2}=0; \quad a_0 \sum_1^n \frac{a_k(\sigma^2+\omega^2+\alpha_k\sigma)}{(\sigma+\alpha_k)^2+\omega^2}=0.$$

A corollary of this theorem, in the case (c) is a straight line or circle, symmetrically placed with respect to the real axis, unifies the three known cases of LC networks, RL–RC networks, and networks with slight dissipation.

Analysis and synthesis of RLC networks possessing a canonic form is introduced in the light of a more general approach to the problem based on the class consideration. The problem of the driving-point impedances which contain mutual coupling in their Brune configuration has been clarified by outlining their generating functions and their network structure. (The latter part is omitted here and shall be presented in another article.)

GENERAL CONSIDERATIONS

LET $z(s)$ be the driving-point impedance of a linear passive network with lumped elements, where $s=\sigma+j\omega$ is the complex frequency, then,

$$z(s)=z(\sigma+j\omega)=U(\sigma,\omega)+jV(\sigma,\omega). \tag{1}$$

Poles and zeros of $z(s)$ are located on the two families of curves:

$$U(\sigma,\omega)=0, \quad \text{curve } (c), \tag{2a}$$

$$V(\sigma,\omega)=0, \quad \text{curve } (r). \tag{2b}$$

Now consider the functional relationship

$$Z(z)=F(z,z^{-1}), \tag{3}$$

presenting any random combination of z and its reciprocal with arbitrary impedance levels such as those shown in Fig. 1. The driving-point impedance or admittance of Fig. 1 has the following form:

$$Z(z)=\frac{Hz(z^2+A_1^2)(z^2+A_2^2)\cdots(z^2+A_n^2)}{(z^2+B_1^2)(z^2+B_2^2)\cdots(z^2+B_m^2)}; \qquad m=\begin{matrix} n \\ \text{or} \\ n+1. \end{matrix} \tag{4}$$

$Z(z)$ may be expanded in the form shown in Eq. (4a).

$$Z(z)=K_0z+\frac{H_0}{z}+\sum_1^m \frac{C_kz}{z^2+B_k^2}. \tag{4a}$$

Poles and zeros of $Z(s)$ are respectively located on curves (5a) and (5b):

$$z\pm jB_k=0. \tag{5a}$$

$$z\pm jA_k=0. \tag{5b}$$

Thus, poles and zeros of $Z(s)$ must be located on the curve (c)

$$\text{Re } z(s)=U(\sigma,\omega)=0. \tag{5}$$

Equation (5) shows that poles and zeros of the driving-point impedance of any finite random combination of z and z^{-1} are located on the same curve (c). The shape of the curve (c) is completely controlled by the electrical characteristics of box z and has nothing to do with the interconnection of the boxes. Figure 2 presents the network configuration of Eq. (4a).

In the particular case of $z=s$, Eq. (5) reduces to Eq. (5c):

$$\text{Re } s=\sigma=0 \tag{5c}$$

which is the well-known class of reactances with poles and zeros alternating on the imaginary axis of the frequency plane.[1]

An analogous example could be given by considering the random combination of a driving-point impedance $z(s)$ with resistors. The functional relationship and the

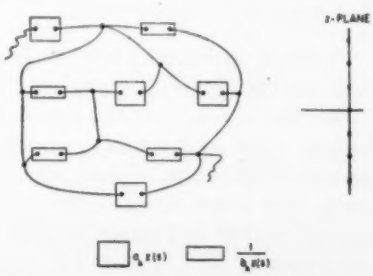


FIG. 1. The random combination of $a_kz(s)$ and $b_kz^{-1}(s)$.

* This work has been supported in part by the U. S. Signal Corps the U. S. Air Materiel Command, and the U. S. Office of Naval Research.

† Presented at the Joint Meeting of the Union Radio Scientifique Internationale and the Institute of Radio Engineers, April 27, 1953, Washington, D. C.

[1] R. M. Foster, Bell System Tech. J. 3, 259–267 (1924).

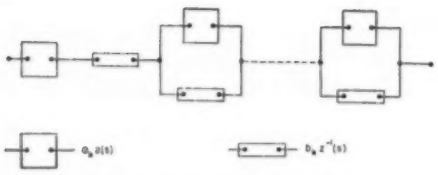


FIG. 2. Canonic representation of the driving-point impedance of Fig. 1.

driving-point impedance $Z(z)$ are given by Eqs. (6) and (7),

$$Z(z)=f(z,R), \tag{6}$$

$$Z(z)=\frac{h(z+a_1)(z+a_2)\cdots(2+a_n)}{(z+b_1)(z+b_2)\cdots(z+b_m)}; \quad m=\begin{matrix}n-1\\ n\\ n+1\end{matrix}. \tag{7}$$

$Z(z)$ may be expanded in the form,

$$Z(z)=h_0+k_0z+\sum_1^m \frac{c_kz}{z+b_k}, \tag{8}$$

or other familiar forms.[2]

Poles and zeros of $Z(z)$ are located on the curve (r):

$$Imz(s)=V(\sigma,\omega)=0. \tag{9}$$

In the particular cases, $z(s)=s$ or $z(s)=s^{-1}$, we have the familiar RL or RC cases of Cauer.[2]

CANONICAL NETWORK CONFIGURATIONS

Foster's expansion of driving-point reactance functions provides canonical configurations for their realization. The reactance function is synthesized into series or parallel combinations of resonant or antiresonant circuits. The question arises whether such a disciplinary form could be found for some class of RLC networks. To answer this question a definition for RLC canonical forms must first be established. As a generalization of Foster's form it seems logical to search for the most general family of driving-impedance functions which could be represented in a Foster-wise structure as shown in Fig. 2.

At first it would seem that any physically realizable driving-point impedance of the form of Eq. (4) should have a canonical structure in the Foster sense; however, further thought reveals the fallacy of this reasoning. In fact, the basis of Foster's reactance form rests on the following two requirements:

A. Poles and zeros of a driving-point reactance alternate on the algebraic curve (c), Re $s=0$.

B. The curve (c) intersects the real axis in a pair of points (zero and infinity) which are also two of the critical frequencies of the reactance function. Furthermore, the symmetrical curve (c) has each of its two halves located in a separate half-plane.

[2] E. A. Guillemin, *Communication Networks* (John Wiley and Sons, Inc., New York, 1946), fifth edition, Vol. II, Chap. V.

DRIVING-POINT IMPEDANCE FUNCTIONS POSSESSING CANONIC CONFIGURATION

Any function representing the random combinations of two finite impedances z and z^{-1}, as shown in Fig. 1, satisfies condition A above. For the curve (c), consider Re $z(s)=0$, which is the generalization of this condition. If $z(s)$ is the generating function which gives rise to a driving-point impedance permitting canonical representation in the Foster sense, then the poles and zeros of $z(s)$ must occur for real values of s.

As an example of the insufficiency of requirement A alone, consider the physically realizable impedance.

$$z(s)=\frac{4s^2+s+2}{s^2+2s+2}.$$

Obviously the function $Z(s)$ does not have real critical frequencies; furthermore, for a pole $z=-jB_k$ of $Z(z)$, one obtains the corresponding two poles for $Z(s)$:

$$(4+jB_k)s^2+(1+2jB_k)s+2(1+jB_k)=0. \tag{10}$$

For finite real values of B_k, one finds that the two roots of Eq. (10) do not lie in the same half-plane; therefore, the function generated by the particular $z(s)$ of the above example does not meet the requirement of condition B.

From the above we conclude that if $Z(s)$ is to possess an RLC canonical structure in Foster's sense it must fulfill the following conditions:

1. $Z(s)$ must be generated by combination of z and z^{-1}.
2. $z(s)$ must have all its poles and zeros real.
3. The half-segment of the curve Re $z(s)=0$ corresponding to $\omega>0$ must not cross the real axis (a similar statement applies for $\omega<0$).

Condition 3 implies that the poles of $Z(s)$ corresponding to a pole $-jB_k$ of $Z(z)$

$$z(s)=f(s)=\frac{p(s)}{q(s)}; \quad p(s)+jB_kq(s)=0 \tag{11}$$

must remain in the upper or lower quadrant of the left half-plane. By the application of Lucas' theorem, one may show that this condition is fulfilled if and only if the poles and zeros of $z(s)$ alternate on the real axis.

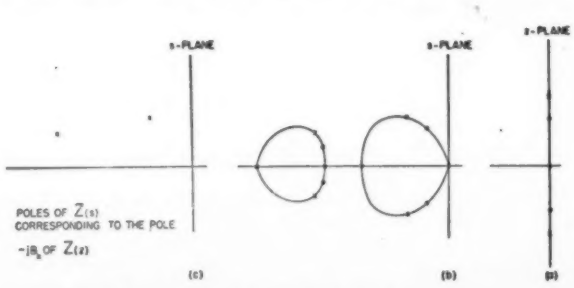


FIG. 3. Generation of canonic impedance function.

(F. M. Reza[3] and M. Marden[4] have discussed more elaborate proofs.) Thus the following theorem is established:

Theorem: A driving-point impedance function can be presented in a canonical form (in the above-mentioned sense) if and only if $Z(s)$ is the composition of the following two functions

$$Z(s)=F[f(s)], \tag{12}$$

when $F(\lambda)$ represents an LC driving-point impedance function, and $f(s)$ an RL or RC function.

Poles and zeros of driving-point impedance functions admitting a canonical form in Foster sense, are located on either of the following two families of curves:

$$f(s)=a_0+\sum_1^n \frac{a_k}{s+\alpha_k} \quad \text{RC case,} \tag{13}$$

$$a_0+\sum_1^n \frac{a_k(\sigma+\alpha_k)}{(\sigma+\alpha_k)^2+\omega^2}=0, \tag{13a}$$

$$f(s)=a_0+\sum_1^m \frac{a_k s}{s+a_k} \quad \text{RL case,} \tag{14}$$

$$a_0+\sum_1^m \frac{a_k(\omega^2+\sigma^2+\alpha_k\sigma)}{(\sigma+\alpha_k)^2+\omega^2}. \tag{14a}$$

The simplest members of the above families are straight lines and circles which pertain to the following interesting cases:

$$A.\ f(s)=\frac{a}{s} \text{ or } as \text{ locus } \sigma=0 \quad \text{reactance family,}$$

$$B.\ F(\lambda)=A_1 \text{ or } \frac{A_1}{\lambda} \text{ locus } \omega=0 \quad \text{RL or RC,}$$

$$C.\ f(s)=\frac{a_1}{s+\alpha_1} \text{ or } a_1(s+\alpha_1) \text{ locus } \sigma=-\alpha \quad \text{RLC,}$$

$$D.\ f(s)=\frac{as}{a+cs} \quad \text{locus circle RLC.}$$

In each of these four cases, it is noticed that poles and zeros alternate on a locus such that some of the critical frequencies[5] coincide with the points where the locus intersects the real axis. This simple remark is of importance since it suggests that for these four types of functions, which possess a canonical form, one can omit the cumbersome computation of determining whether $\text{Re}\, Z(s)\gtrless 0$ for $\text{Re}\, s\gtrless 0$. Thus the following is established as a corollary of the general theorem.

Corollary: If poles and zeros of a real rational function with a positive constant multiplier alternate on a circle or straight line, symmetrically located with respect to the real axis in the left half-plane or its boundary, the function will present a driving-point impedance with a canonic form.

This corollary consolidates the following three known statements:

A. For LC case, take the straight line as the imaginary axis.[1]

B. For RC—RL case, take the straight line as the real axis.[2,6]

C. For a particular case of RLC "slightly lossy networks," discussed by E. A. Guillemin and J. G. Linvill,[7] take the straight line parallel to the imaginary axis.

It is noticed that in the general case, locus (c) has several "separate parts."[8] The alternance property of poles and zeros on each and every "separate part" is a necessary but not sufficient condition for physical realizability. A given set of critical frequencies of $Z(s)$ alternating on a "part" of (c) completely determines the corresponding critical frequencies for every other "separate part" of (c).

CLASS OF A CANONICAL DRIVING-POINT IMPEDANCE FUNCTION

Assume the curve (c) obtained from $\text{Re}\, Z(s)=0$ to be a "generalized Foster locus" described by Eq. (13a) or (14a). We may conveniently define the driving point of any interconnection of $Z(s)$ and $Z^{-1}(s)$ as an impedance of the canonical class (c). From this definition we conclude:

If $Z_1, Z_2, \cdots$, and Z_n are of the same canonical class, then any finite network constructed by an interconnection of them will present a driving-point impedance $\psi(Z_1, Z_2, \cdots, Z_n)$ which is also of the same class (c).

The above may be verified by considering the transformation $\lambda=f(s)$ which transforms $Z_1, Z_2, \cdots, Z_n$ into $F_1(\lambda), F_2(\lambda)$, respectively. As far as the λ plane is concerned, $F_k(\lambda)$ is a reactance function; therefore $\psi(F_1, F_2, \cdots, F_n)$ is transformable into $\psi[f(s)]$, which represents an impedance of the canonical class (c).

The consideration of the class of networks is not purely theoretical. It could be applied in engineering applications as an extension of the use of the lossless

[3] F. M. Reza, Quarterly Progress Reports (Research Laboratory of Electronics, Massachusetts Institute of Technology), October 15, 1952, pp. 55–59, and April 15, 1953, pp. 57–61.

[4] M. Marden, *Geometry of Poles and Zeros of a Polynomial in a Complex Variable* (Am. Math. Soc., New York, 1949).

[5] The expression "critical frequencies" in this paper refers to poles and zeros of the impedance or admittance functions. This is in conformity with the practice of electrical engineers and should not be confused with "critical points" as used in mathematics.

[6] Though the statement is correct, for more harmony one should note that in this limiting case the locus 14 or 14a merely cuts the negative real axis at the points which are the only critical points of the impedance function. The real axis and the locus do not coincide. In fact, the negative real axis belongs to the family of Eq. (9).

[7] E. A. Guillemin and J. G. Linvill, Quarterly Progress Report (Research Laboratory of Electronics, Massachusetts Institute of Technology), April 15, 1951, p. 57.

[8] The nontechnical word "part" here is used to denote a segment of the curve (c) starting from a point on the real axis and terminating in the succeeding real point of the curve; see example 1 for locus with one part, and example 2 for locus with two parts.

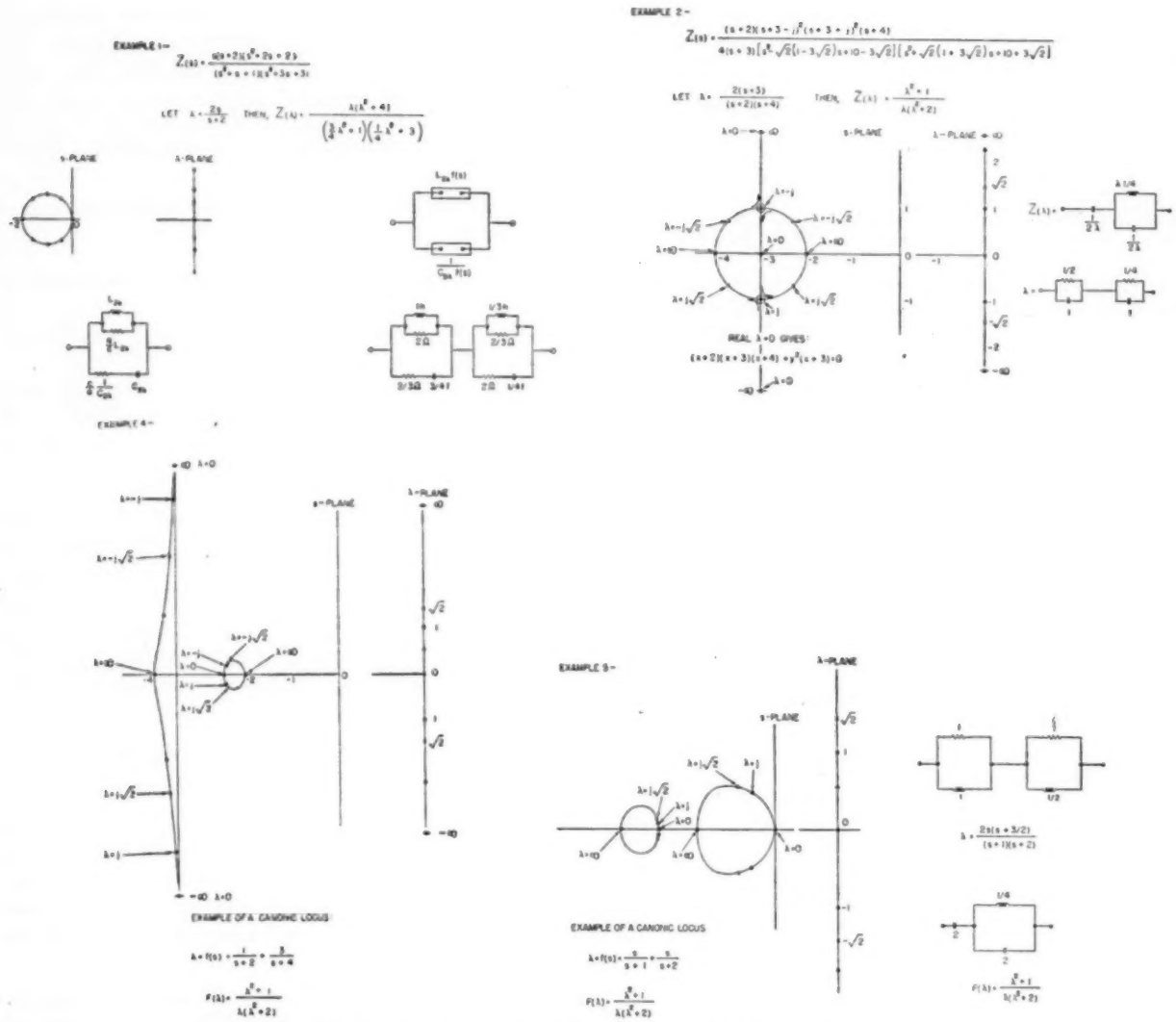


FIG. 4. Examples of synthesis of RLC canonical driving-point impedances.

circuits. First of all, the synthesis of networks with canonical structure is highly facilitated by the above consideration. Secondly, the problem of the analysis is also much more under control. For instance suppose for certain engineering application one is dealing with networks of canonical class (c). The locus (c) in the s plane contains all the poles and zeros of all the functions which are members of this class; therefore the critical frequencies of all these functions are confined to a particular region of the frequency plane. Also since the electrical characteristics of the network depend mainly on its critical frequencies, it is natural to expect that the energy consideration and transient studies may be further simplified by a general study of the relevant characteristics of the considered class.

SYNTHESIS OF RLC NETWORKS POSSESSING CANONIC FORM

An impedance function with canonic form must be the result of the composition of two functions $f(s)$ and $F(\lambda)$ as presented in the principal theorem mentioned above. All critical frequencies of $f(s)$ being real, $Z(s)$ can have one of the two following forms:

A. The real critical frequencies of $Z(s)$ are either all poles or all zeros. Then $Z(s)$ contains a factor

$$\prod_{1}^{k=2n} (s+s_k)$$

in the numerator or the denominator with all s_k positive real numbers.

B. Some of the real critical frequencies of $Z(s)$ are poles and some zeros. Then $Z(s)$ contains a factor

$$\frac{\prod_{1}^{k=n_1} (s+s_k)}{\prod_{1}^{j=n_2} (s+s_j)},$$

where s_k and s_j are distinct real positive numbers.

Case A

First, one may write down

$$\frac{\prod_{1}^{k=n_1}(s+s_k)}{\prod_{1}^{j=n_2}(s+s_j)}=\lambda(s) \qquad (15)$$

and make sure that the poles and zeros of $\lambda(s)$ alternate on the negative real axis. Then if $Z(s)$ possesses a canonic form, it must belong to the class represented by $\mathrm{Re}\,[\lambda(s)]=0$ of Eq. (15). Therefore, poles and zeros of $Z(s)$ must alternate on each and every "part" of the locus $\mathrm{Re}\,\lambda=0$. The transformation, Eq. (15), will transform $Z(s)$ into a reactance $Z(\lambda)$ (see Example 2).

Case B

One writes down the ratio

$$\frac{\prod_{1}^{k=n_1}(s+s_k)}{\prod_{1}^{j=n_2}(s+s_j)}=\lambda(s),$$

where the roots of its numerator and denominator are selected alternatively among the succeeding real critical frequencies of $Z(s)$. If $Z(s)$ possesses a canonic form, then a transformation similar to Eq. (15) will transform it into a reactance (see Example 1).

EXAMPLES OF SYNTHESIS

Example 1

$$z(s)=\frac{s(s+2)(s^2+2s+2)}{(s^2+s+1)(s^2+3s+3)}.$$

If $Z(s)$ has a canonic form, then

$$\lambda=k(s/s+2)$$

(k is positive constant) must determine its class. In fact, $\mathrm{Re}\,\lambda=0$ gives

$$(\sigma+1)^2+\omega^2=1,$$

which represents a circle in the left half-plane. Poles and zeros of $Z(s)$ alternate on this circle; therefore, in conformity with the corollary of our general theorem, $Z(s)$ could present a physically realizable driving-point impedance in canonic form (see Fig. 4, Example 1).

Example 2

$$z(s)=\frac{(s+2)(s+3-j)^2(s+3+j)^2(s+4)}{(s+3)[s^2-\sqrt{2}(1-3\sqrt{2})s+10-3\sqrt{2}][s^2+\sqrt{2}(1+3\sqrt{2})s+10+3\sqrt{2}]}.$$

According to aforementioned rule, let

$$\lambda=k\frac{s+3}{(s+2)(s+4)}.$$

$\mathrm{Re}\,\lambda=0$ presents a third-degree curve which degenerates into a circle and a straight line, $(\sigma+3)[\sigma^2+\omega^2+6\sigma+8]=0$. The curve has two parts in the sense discussed before. As we move along the negative segment of the axis of imaginaries from minus infinity toward zero, the corresponding points of the s plane start moving on the locus from the points $s=-2$ and $s=-4$, as shown in Fig. 4, Example 2.

Note that $\lambda=p(s)/q(s)$ and $1/\lambda$ are both of the same class; therefore, in the selection of λ, one can choose either of these forms. Also the selection of the proper k is not of importance for our present studies.

Example 3

The generating impedance function λ is given as

$$z(s)=\frac{1+3s+2s^2}{1+7s+12s^2}.$$

Poles and zeros are all real, but the alternance on the real axis is not preserved; the functions generated by this $z(s)$ are not realizable in an RLC canonical structure.

Example 4

Study the network representation of $F[\lambda(s)]$ where

$$F(\lambda)=\frac{\lambda^2+1}{\lambda(\lambda^2+2)} \quad \lambda(s)=\frac{s}{s+1}+\frac{s}{s+2}.$$

Re (λ) gives

$$\sigma(\sigma+1)(\sigma+2)^2+\sigma(\sigma+2)(\sigma+1)^2+\omega^2\{[\sigma(\sigma+1)+(\sigma+2)^2]+[\sigma(\sigma+2)+(\sigma+1)^2]\}+2\omega^4=0.$$

Poles and zeros alternate on the two "parts" of the locus as shown in Fig. 4, Example 4.

Example 5

$F[\lambda(s)]$, where

$$F(\lambda)=\frac{\lambda^2+1}{\lambda(\lambda^2+2)} \quad \lambda(s)=\frac{1}{s+2}+\frac{3}{s+4}$$

(see Fig. 4, Example 5).

ACKNOWLEDGMENT

The author wishes to express his gratitude to Professor E. A. Guillemin and Professor J. B. Wiesner for their guidance and encouragement, and to Professor R. B. Adler for his invaluable suggestions and criticisms.

A Resonant Cavity Study of Semiconductors*

HSI-TEH HSIEH, JAMES M. GOLDEY, AND SANBORN C. BROWN
Solid-State and Molecular Theory Group, and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received June 18, 1953)

A solution of Maxwell's equations is obtained in a resonant cavity with a center post of arbitrary electrical properties. The solution gives the dielectric coefficient and the conductivity of the center post in terms of the natural frequency and Q of the cavity. The theory is of particular use in the study of semiconductors where perturbation theories are of little value. It is shown that a transition from a cylindrical mode to a coaxial mode occurs as the conductivity of the center post is varied. This transition occurs for a relatively small change in conductivity. The present results are compared with those of perturbation theory, and it is shown that the latter are valid over a greater range than the conditions imposed in their derivation indicate.

TABLE I. Table of symbols.

a	cavity radius
b	sample radius
μ_0	permeability of free space, 1.257×10^{-6} henry/meter
ϵ_0	permittivity of free space, 8.854×10^{-12} farad/meter
ϵ	inherent permittivity of the medium due to bound charges
$\epsilon_c=\epsilon_r-j\epsilon_j$	permittivity of the medium due to all the charges
ϵ_r	real part of the permittivity: $\epsilon_r=\epsilon+\sigma_j/\omega_s$
ϵ_j	imaginary part of the permittivity: $\epsilon_j=\sigma_r/\omega_s$
ϵ/ϵ_0	inherent dielectric coefficient of the medium
ϵ_c'/ϵ_0	total dielectric coefficient
$\epsilon_r/\epsilon_0=K$	real part of the total dielectric coefficient
$\epsilon_j/\epsilon_0=K\tan\delta$	imaginary part of the total dielectric coefficient
$\epsilon_j/\epsilon_r=\tan\delta$	loss tangent
$\sigma=\sigma_r+j\sigma_j$	complex conductivity of the medium
σ_r	in-phase component of the conductivity
σ_j	out-of-phase component of the conductivity
ω_0	resonant and natural frequency of a cavity without losses
$\omega_t=\omega_r+j\omega_j$	complex natural frequency of cavity with losses
ω_r	real part of the complex frequency
ω_j	imaginary part of the complex frequency
ω_s	resonant frequency of the cavity for forced oscillation
$k=k_r+jk_j$	complex wave number external to the sample
k_r	real part of k
k_j	imaginary part of k
k'	complex wave number internal to the sample
ϕ	phase of the complex frequency and external wave number: $\omega_t=\lvert\omega_t\rvert\exp(j\phi)$
Q	ratio of energy stored to energy dissipated per radian in circuit
Q_0	unloaded Q for cavity with no sample
Q_1	unloaded Q for cavity with sample
Q_e	represents loss in the sample: $(1/Q_e)=(1/Q_1)-(1/Q_0)$, related to ϕ by the relation $Q_e=\frac{1}{2}\cot\phi$

THE solutions of Maxwell's equations for a cavity resonator can be expressed as a linear combination of normal modes. The various modes of oscillation depend upon the geometry and material of which the resonant cavity is made. In practice a cavity is designed to operate in a particular mode and the resonant frequency as a function of dimensions of ideal cavities has been published by various authors.[1]

* This work has been supported in part by the U. S. Army Signal Corps, the U. S. Air Materiel Command, and the U. S. Office of Naval Research.

[1] See, for example, C. G. Montgomery, *Techniques of Microwave Measurements* (McGraw-Hill Book Company, Inc., 1947), Radiation Laboratory Technical Series, Vol. 11, pp. 293–308.

In an oscillating system three frequencies are of interest. The resonant frequency, ω_0, corresponding to undamped oscillations is changed slightly when losses are present. In a cavity in which the field is decaying, the natural frequency is given by $\omega_r=\omega_0[1-(1/2Q)^2]^{\frac{1}{2}}$. If the cavity is in forced oscillation, resonance occurs at some other frequency ω_s, which is related to ω_0 and ω_r by some function of Q. The quantity Q is defined as the ratio of the energy stored to the energy dissipated in the cavity per radian.

We shall treat the transient and the steady-state solutions separately. First we shall obtain the transient solution, and then show how it may be modified to fit the steady-state case.

TRANSIENT SOLUTION

If one considers the case of the field decaying in a resonator, it follows from the definition of Q that the energy decreases as $\exp(-\omega_r/Q)t$. Hence the field will vary as $\exp j(\omega_r+j\omega_r/2Q)t$, which may be written as $\exp j(\omega_r+j\omega_j)t$. Thus when losses are present, the frequency $\omega_t=\omega_r+j\omega_j$ is complex and we have the relations

$$Q=\tfrac{1}{2}(\omega_r/\omega_j)=\tfrac{1}{2}\cot\phi;\quad \phi=\tan^{-1}(\omega_j/\omega_r). \qquad (1)$$

Thus we can obtain Q quite simply from ϕ, the angle of the complex frequency.

It is important to realize that the concept of a complex frequency is valid only for the case of energy decay. In practice one drives the cavity by a magnetron or klystron and there is no energy decay; the energy which is dissipated is replaced by the generator. The solution obtained here is for the transient case and one must be careful in applying the results of this section to steady-state measurements when low Q's characterize the resonance.

In the present paper we consider a cylindrical cavity resonator oscillating in the TM_{010} mode and investigate the results of inserting samples of various conductivity and dielectric coefficient as axial posts in the cavity. We may determine the dielectric coefficient and conductivity by measuring Q and the resonant frequency. It is convenient to use the product of a complex wave

number k and the cavity radius a as a characteristic value, since this product is a dimensionless variable, rather than using the actual frequency. For an empty cylindrical cavity in the TM_{010} mode, ka, which may be considered a real quantity in this case, is equal to 2.4048. If we make the resonator a coaxial cavity by inserting an axial metal post of radius b, and take as an example $a/b=100$, then $ka=2.801$, where again this may be considered a real number. The interesting problem arises when we introduce an axial dielectric or semiconducting post.

We shall employ the concept of a complex electric permittivity $\epsilon_e=\epsilon_r-j\epsilon_j$ in describing the medium. The inherent permittivity of the medium is denoted by ϵ, and the inherent dielectric coefficient is given by ϵ/ϵ_0.

In an alternating field, the conductivity of a medium is a complex quantity, $\sigma=\sigma_r+j\sigma_j$. The imaginary part, σ_j, represents the motion of the electrons which is out of phase with the field. This is usually a negative quantity.

In order to determine the physical properties represented by the complex permittivity we consider the total conductivity, arising from both displacement and conduction currents in the medium. In the steady state it is given by

$$j\omega_s\epsilon+\sigma=j\omega_s(\epsilon-j\sigma/\omega_s). \tag{2}$$

In terms of a complex permittivity this may be written as

$$j\omega_s(\epsilon_r-j\epsilon_j). \tag{3}$$

When we use a complex expression for the conductivity in Eq. (2) the term in parentheses becomes

$$\epsilon-j(\sigma_r+j\sigma_j)/\omega_s=\epsilon+\sigma_j/\omega_s-j\frac{\sigma_r}{\omega_s}. \tag{4}$$

We assume that ϵ itself is real, all losses being attributed to the conductivity of the medium. Comparing Eq. (4) with Eq. (3) we have

$$\epsilon_r=\epsilon+\sigma_j/\omega_s \tag{5}$$

and

$$\epsilon_j=\sigma_r/\omega_s. \tag{6}$$

We may define two quantities, K and $\tan\delta$, as follows: $K=\epsilon_r/\epsilon_0$ and $\tan\delta=\epsilon_j/\epsilon_r$. The quantity K is equal to the inherent dielectric coefficient of the medium in the low conductivity limit, where $\epsilon_r=\epsilon$, and the loss tangent, $\tan\delta$, is a measure of the loss in the medium. When ϵ is real, as it is assumed to be here, and is considerably larger than the quantity σ_j/ω_s, $\tan\delta$ is essentially equal to $\sigma_r/\omega\epsilon$. The size of this quantity relative to unity can be used to distinguish between a dielectric and a conducting medium. If $\sigma_r/\omega\epsilon$ is very much less than unity, the medium is dielectric, whereas when it is greater than about 10^5 the medium is conducting. A transition from dielectric to conducting properties occurs in the intermediate region.

Bethe and Schwinger[2] treated the problem of a dielectric post by a perturbation method and obtained the following expressions which are valid under the restrictions of small losses for any mode where the electric field is parallel to the sample surface.

$$\frac{\lambda_1-\lambda_0}{\lambda_0}=\frac{\left(\frac{\epsilon_r}{\epsilon_0}-1\right)\int_{\text{sample}}E^2dv}{2\int_{\text{cavity}}E^2dv}, \tag{7}$$

$$\frac{1}{Q_e}=\frac{1}{Q_1}-\frac{1}{Q_0}=\frac{\epsilon_j}{\epsilon_0}\frac{\int_{\text{sample}}E^2dv}{\int_{\text{cavity}}E^2dv}, \tag{8}$$

where λ_0 and λ_1 are the resonant wavelengths of the empty and loaded cavity, respectively. Q_0 and Q_1 are the unloaded Q's of the empty cavity and the cavity containing an axial post, and Q_e is a measure of the loss in the sample.

An exact solution may be obtained by finding solutions of the vector wave equation, both external, Eq. (9), and internal, Eq. (10), to the axial post:

$$\nabla\times\nabla\times\mathbf{E}-k^2\mathbf{E}=0, \tag{9}$$

$$\nabla\times\nabla\times\mathbf{E}-k'^2\mathbf{E}=0, \tag{10}$$

where

$$k=(\mu_0\epsilon_0)^{\frac{1}{2}}\omega_t \tag{11}$$

and

$$k'=[\mu_0(\epsilon_r-j\epsilon_j)]^{\frac{1}{2}}\omega_t=[(\epsilon_r/\epsilon_0)(1-j\epsilon_j/\epsilon_r)]^{\frac{1}{2}}k. \tag{12}$$

A general solution is difficult, but one can be evaluated for the TM_{010} mode. In this mode the electric field has an axial component only and the magnetic field has only an angular component. The solutions of Eq. (9) outside the sample are

$$E=\beta[J_0(kr)-\alpha N_0(kr)], \tag{13}$$

$$H=j\beta(\epsilon_0/\mu_0)^{\frac{1}{2}}[J_1(kr)-\alpha N_1(kr)]. \tag{14}$$

We require the electric field to vanish at the cavity wall. This determines the constant α and we have

$$E=\beta[J_0(kr)-J_0(ka)N_0(kr)/N_0(ka)], \tag{15}$$

$$H=j\beta(\epsilon_0/\mu_0)^{\frac{1}{2}}[J_1(kr)-J_0(ka)N_1(kr)/N_0(ka)]. \tag{16}$$

The solutions of Eq. (10) within the axial post are

$$E=\xi J_0(k'r), \tag{17}$$

$$H=j\xi(\epsilon_0/\mu_0)^{\frac{1}{2}}(k'/k)J_1(k'r). \tag{18}$$

E and H must be continuous at the surface of the post, $r=b$. From this condition we obtain, by matching the

[2] H. A. Bethe and J. Schwinger, National Defense Research Committee Report D1-117 (1943); see also G. Birnbaum and J. Franeau, J. Appl. Phys. **20**, 817 (1949).

ratio of H to E at the surface of the axial post, the following transcendental equation:

$$k'\frac{J_1(k'b)}{J_0(k'b)}=k\frac{J_1(kb)N_0(ka)-J_0(ka)N_1(kb)}{J_0(kb)N_0(ka)-J_0(ka)N_0(kb)}. \quad (19)$$

When small axial posts are used, approximate formulas for the Bessel and Neumann functions may be used.

Numerical solutions have been obtained in the present work using both real and imaginary terms of Eq. (19). However, we have also made a power series expansion of the complex terms in Eq. (19), and find the result may be expressed by the following formulas:

$$\epsilon_r/\epsilon_0=1+A+B\phi^2, \quad (20)$$

$$\epsilon_j/\epsilon_0=C\phi, \quad (21)$$

where A, B, and C are given by

$$A=\frac{\pi J_0(D)N_0(D)+2\ln(2/\gamma F)J_0^2(D)}{F^2[(\pi/2)N_0(D)+\ln(2/\gamma F)J_0(D)]^2},$$

$$B=\frac{\pi D^2J_1(D)N_1(D)+2D^2\ln(2/\gamma F)J_1^2(D)+2\pi D[J_0(D)N_1(D)-J_1(D)N_0(D)]}{F^2[(\pi/2)N_0(D)+\ln(2/\gamma F)J_0(D)]^2},$$

$$C=\frac{\pi D[J_1(D)N_0(D)-J_0(D)N_1(D)]+2\pi J_0(D)N_0(D)}{F^2[(\pi/2)N_0(D)+\ln(2/\gamma F)J_0(D)]^2},$$

$$D=|ka|\,;\quad F=|kb|\,;\quad \gamma=1.7811.$$

The angle ϕ is to be calculated from Q_e in accordance with Eq. (1). The results of this calculation are compared with the more exact solution obtained below in Figs. 4 and 5. The expansion is only valid in the negative frequency shift region.

Feenberg[3] has also made an expansion of Eq. (19), neglecting all imaginary terms. His result for a small axial post is $\epsilon_r/\epsilon_0=1+0.224(a/b)^2(2.4048-k_ra)$.

Numerical solutions obtained in the present work using both real and imaginary terms of Eq. (19) are shown in the following figures. For purposes of illustration we have restricted our calculations to a ratio of cavity to axial post radius of 100.[4] Furthermore the curves for (ϵ_r/ϵ_0) are calculated in the neighborhood of 16 for its applicability to the study of germanium.

For any given ratio of resonant cavity radius to axial post radius there are four variables of interest. Two of these, the dielectric coefficient and the conductivity of the axial post, are properties of the material of the post; and two are measured quantities of the resonator, the Q and the resonant frequency.

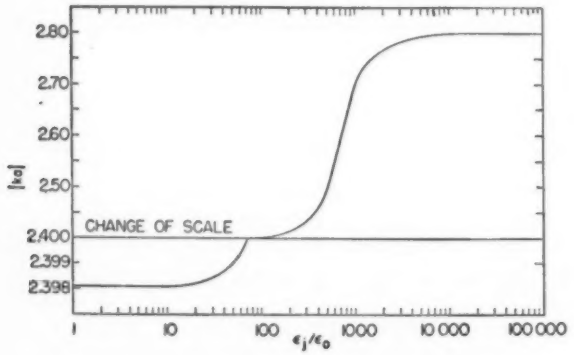


FIG. 1. Magnitude of ka as a function of (ϵ_j/ϵ_0) for $a/b=100$ and $(\epsilon_r/\epsilon_0)=16$.

In Fig. 1 is plotted the magnitude of the product of the characteristic resonant wave number and the cavity radius ka as a function of (ϵ_j/ϵ_0) when (ϵ_r/ϵ_0) is 16. The transition from the cylindrical resonant mode to the coaxial mode occurs for this case when (ϵ_j/ϵ_0) varies from 10.5 to about 8000. This corresponds in the 10-cm wavelength region to a variation in conductivity from 0.018 mhos per cm to 15.6 mhos per cm. An initial decrease in frequency occurs when a low-conductivity axial post is introduced. As the conductivity is increased, however, the frequency increases slowly until it passes through the empty cavity resonance. Further increase in conductivity causes a more rapid frequency rise, the frequency asymptotically approaching the value for a coaxial resonator.

The presence of bound electrons, which give rise to a polarization of the sample when an external field is applied, causes a frequency decrease in the cavity resonator. Conduction electrons, made available by increasing the conductivity of the axial post, tend to cause higher resonant frequencies. The opposing effects of bound and conduction electrons neutralize each other for $(\epsilon_j/\epsilon_0)=130$. Here the frequency shift referred to the empty cylindrical cavity is zero. As more conduction electrons are added, they continue to dominate the bound ones and the coaxial limit is reached for all practical purposes for the case illustrated here when their density reaches 2.7×10^{16} electrons per cc in germanium.[5]

In Fig. 2 we show a plot of Q_e as a function of (ϵ_j/ϵ_0),

[3] E. Feenberg, Sperry Gyroscope Company Report GC16907, July, 1942.

[4] More complete calculations are detailed in Technical Report 2, by Hsi-Teh Hsieh, Solid State and Molecular Theory Group, Massachusetts Institute of Technology, Map 1, 1952.

[5] Using an electron mobility of 3600 cm²/volt-sec. E. M. Conwell, Proc. Inst. Radio Engrs. 40, 1327 (1952).

again for $(\epsilon_r/\epsilon_0)=16$. The change in Q_e reflects two opposing effects, in this case both attributable to conduction electrons. The introduction of these electrons into the ac field causes ordinary ohmic losses. However, they also shield the interior of the post. The dissipative effect of the electrons is dominant initially, causing the Q_e to drop to 5.2. Thereafter the shielding effect dominates and Q_e increases, approaching infinity with the conductivity.

Figure 3 is a plot of the complex permittivity, relative to that of free space, in the complex plane of the variable ka. The quantity $k_r a$ is related to the natural frequency through Eq. (11) and $k_j a$ is related to Q_e through Eqs. (11) and (1). The contours of constant (ϵ_r/ϵ_0) differ from each other at the low-conductivity end by the amount given by the perturbation theory. At the coaxial limit the curves meet, for here only the conduction electrons are important. The constant (ϵ_j/ϵ_0) curves all converge toward the point $(2.801+j0)$ but do not actually reach this point. This is because the imaginary part of ka cannot be zero (an infinite Q)

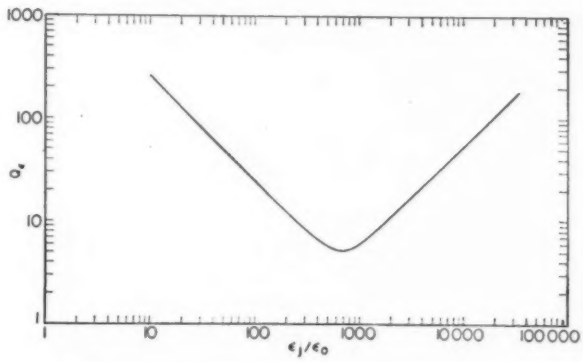


FIG. 2. Q_e as a function of (ϵ_j/ϵ_0) for $a/b=100$ and $(\epsilon_r/\epsilon_0)=16$.

for a finite conductivity. The point, however, does represent the complete contour for the limiting case of $(\epsilon_j/\epsilon_0)=\infty$. The other limiting case, $(\epsilon_j/\epsilon_0)=0$, is the positive real axis below 2.4048. On this axis, (ϵ_r/ϵ_0) is equal to the inherent dielectric coefficient, as mentioned above.

Since the properties of the center post uniquely determine the Q_e and the resonant frequency it is, in principle, possible to determine these properties from measurement of Q_e and ω_t. However since the Q drops to very low values in the region of positive frequency shifts it is desirable from an experimental point of view to confine measurements to the negative shift region. Figures 4 and 5 may be used to determine (ϵ_r/ϵ_0) and (ϵ_j/ϵ_0) from measured frequency shifts and Q's in this region.

In Fig. 4 (ϵ_r/ϵ_0) is plotted as a function of ϕ with the magnitude of ka as a parameter. For purposes of both calculation and plotting, ϕ is a more convenient quantity to use than Q_e. The curves for the various values of ka are equally spaced, hence a linear interpolation

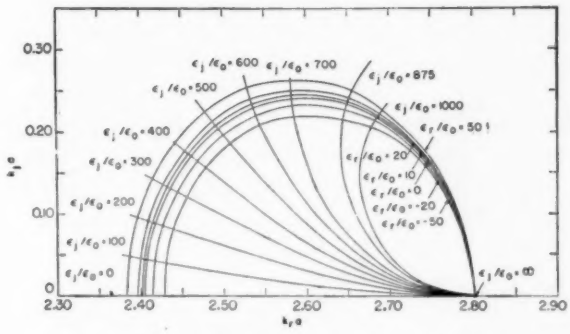


FIG. 3. Contours of constant (ϵ_r/ϵ_0) and (ϵ_j/ϵ_0) in the complex ka plane for $a/b=100$.

may be used for those values not given. In Fig. 5 the quantity (ϵ_j/ϵ_0) is plotted as a function of ϕ with ka again as the parameter. It is to be noted that these curves are practically coincident. This means that in this range the Q_e is determined primarily by the dissipation in the material. The frequency shift, on the other hand, as seen in Fig. 4, is determined by both dielectric and dissipative factors.

Both Figs. 1 and 3 show that the earlier results of both Bethe and Schwinger and of Feenberg may be used with small error up to a value of (ϵ_j/ϵ_0) of about 10, corresponding to a value of 0.6 for (ϵ_j/ϵ_r) in the present case. Similarly Fig. 4 shows that they are valid when $a/b=100$ for ϕ as large as 0.05°, which corresponds to a Q_e as low as 600. Thus the restrictions used in obtaining those results, $(\epsilon_j/\epsilon_r)\ll 1$, are actually overly stringent. The results of the simpler theories may be satisfactorily used for relatively lossy substances.

STEADY-STATE SOLUTION

Losses in germanium may be high and as a result the cavity Q may be low. When the value of Q is low, the steady-state resonant frequency may differ appreciably from the natural decay frequency.

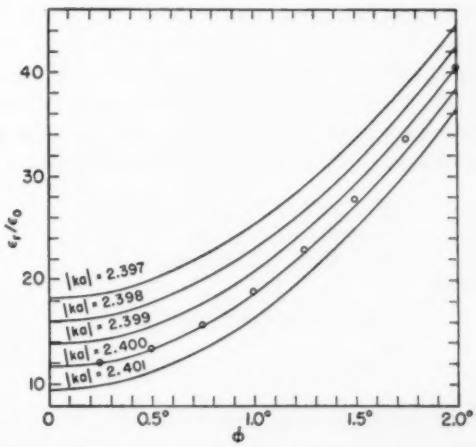


FIG. 4. (ϵ_r/ϵ_0) as a function of ϕ for various values of ka for $a/b=100$. The solid lines represent the numerical solution of Eq. (19) and the points give (ϵ_r/ϵ_0) as calculated from Eq. (20) for $|ka|=2.4000$.

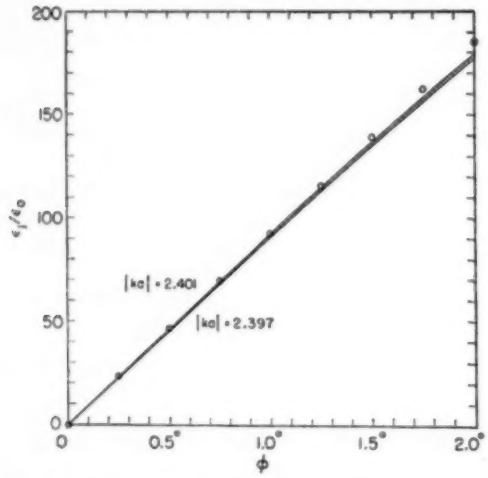


FIG. 5. (ϵ_j/ϵ_0) as a function of ϕ for several values of ka when $a/b=100$. The solid lines represent the numerical solution of Eq. (19) and the points give (ϵ_j/ϵ_0) as calculated from Eq. (21) for $|ka|=2.401$.

A steady-state oscillation is different in several respects from a transient decay. In the steady-state case the frequency is real. In addition, whenever losses are present, the frequency of oscillation may assume any value, not merely certain discrete values as in the transient case. Finally, one must account for the power input in a steady-state oscillation when losses are present.

Although it is possible in principle to obtain a solution to Maxwell's equations in a resonant cavity containing loops, irises, or other coupling devices, it is not possible in practice to do so. We approximate the physical situation by assuming that the power flows into the cavity uniformly and radially through the walls. Mathematically this means that we no longer require the electric field to vanish at the cavity walls. One is then able to obtain a solution for any real value of the frequency ω.

The fields inside the sample are given by Eqs. (17) and (18) and those outside by Eqs. (13) and (14). The quantities k and k' are defined as in Eqs. (11) and (12) except that ω_t is now replaced by ω. At resonance $\omega=\omega_s$. The constant appearing in the external solution is to be determined. The ratio of H and E at the sample surface $(r=b)$ is again matched and we obtain

$$\frac{k'}{k}\frac{J_1(k'b)}{J_0(k'b)}=\frac{J_1(kb)-\alpha N_1(kb)}{J_0(kb)-\alpha N_0(kb)}. \tag{22}$$

We define

$$Z_1(kr)\equiv J_1(kr)-\alpha N_1(kr)$$

and

$$Z_0(kr)\equiv J_0(kr)-\alpha N_0(kr).$$

Denoting the ratio of $Z_1(kb)$ to $Z_0(kb)$ by η, Eq. (22) becomes

$$\frac{k'}{k}\frac{J_1(k'b)}{J_0(k'b)}=\frac{Z_1(kb)}{Z_0(kb)}=\eta=\eta_r+j\eta_j. \tag{23}$$

From this expression one can calculate α for a given set of parameters.

The criterion of resonance for this case is the equality of stored electric and magnetic energy. The electric and magnetic energies are calculated from Eqs. (11)–(14), and one obtains

$$W_M=\epsilon_0\pi a^2\Bigg\{Z_1(D)Z_1^*(D)+Z_0(D)Z_0^*(D)$$
$$-\frac{2}{D}\,\mathrm{Re}Z_0(D)Z_1^*(D)+Z_0(F)Z_0^*(F)$$
$$\times\left(\frac{b^2}{a^2}\right)\Bigg\{-1+\frac{\epsilon_r}{\epsilon_0}\left(1+\frac{\epsilon_j^2}{\epsilon_r^2}\right)^{\frac{1}{2}}$$
$$+\frac{2}{F}\,\mathrm{Re}\left[\eta\left(1-\exp j\tan^{-1}\frac{\epsilon_j}{\epsilon_r}\right)\right]\Bigg\}\Bigg\}, \tag{24}$$

$$W_E=\epsilon_0\pi a^2\Bigg\{Z_1(D)Z_1^*(D)+Z_0(D)Z_0^*(D)$$
$$+\left(\frac{b^2}{a^2}\right)Z_0(F)Z_0^*(F)\left[\left(\frac{\epsilon_r}{\epsilon_0}\right)-1\right.$$
$$\left.-\eta\eta^*\left(1-\frac{\epsilon_r}{(\epsilon_r^2+\epsilon_j^2)^{\frac{1}{2}}}\right)\right]\Bigg\}. \tag{25}$$

In these equations W_M and W_E are the magnetic and electric energies, respectively. Re denotes the real part of the expressions following: D is ka and F is kb.

When the condition of resonance $(W_M-W_E)/\epsilon_0\pi a^2=0$ is applied, one obtains

$$\frac{a}{b}\frac{\mathrm{Re}Z_0(D)Z_1^*(D)}{Z_0(F)Z_0^*(F)}=\mathrm{Re}\left[\eta\left(1-\exp j\tan^{-1}\frac{\epsilon_j}{\epsilon_r}\right)\right]$$
$$+\frac{F}{2}\left[\left(1+\frac{\epsilon_j^2}{\epsilon_r^2}\right)^{\frac{1}{2}}-1\right]\left[\frac{\epsilon_r}{\epsilon_0}+\frac{\eta\eta^*}{\left(1+\frac{\epsilon_j^2}{\epsilon_r^2}\right)^{\frac{1}{2}}}\right]. \tag{26}$$

The last equation can be solved only by a trial and error method. Hence it is not of great practical use in itself However, we have obtained solutions for various values of the parameters and compared the results with those obtained in the transient case. We find, by purely empirical methods, that there is a definite relation for a certain range of Q_e between the resonant frequency in this case and the natural decay frequency obtained

above. The two frequencies are related as follows:

$$\omega_r=\left[1-\left(\frac{5}{4Q_e^2}\right)\right]^{\frac{1}{2}}\omega_s. \quad (27)$$

Steady-state resonant frequencies can be converted to ω_r by this relation, and then the previous solution may be used. The relation is good for values of Q_e lying between ∞ and 30. For frequency measurements to five significant figures it is necessary to correct ω_s for Q's lower than 300.

The authors are indebted to Professor W. P. Allis for many helpful ideas and discussion on the problems involved in this study.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 3 MARCH, 1954

16 Percent Aluminum-Iron Alloy Cold Rolled in the Order-Disorder Temperature Range*

JOSEPH F. NACHMAN AND WILLIAM J. BUEHLER
U. S. Naval Ordnance Laboratory, White Oak, Silver Spring, Maryland
(Received July 6, 1953)

The methods of fabricating 16 percent Al-Fe from cast slab to thin-gauge sheet are described in some detail. The melting, casting, homogenizing, hot rolling, and cold rolling at 575°C and room temperature are described. Particular attention is focused upon the 575°C cold rolling from the standpoint of the possible beneficial effects derived from an ordering reaction which occurs in this alloy. The technique of cold reduction from 0.007 in. to 0.0005 in. at room temperature is discussed.

Magnetic data on a limited number of heat-treated laminated cores are given. Useful physical properties, other than magnetic, of previous limited interest because of the inability to fabricate the alloy into ductile strong thin sheets, are discussed. These properties include excellent oxidation resistance at high temperatures, good wet-corrosion resistance to certain chemical solutions, and high electrical resistivity.

INTRODUCTION

THE development of a process for cold working brittle alloys of iron and aluminum resulted from efforts to produce high-quality magnetic materials containing no alloying elements in critical supply. The need for conservation of cobalt and nickel commonly used in high-quality magnetic materials was a prime consideration.

Extensive research has been done in this country on Si-Fe magnetic alloys. The Al-Fe alloys, which also have very interesting magnetic properties, have received less attention because of the extreme difficulty experienced in fabrication due to their inherent hard and brittle characteristics. This is particularly true of the higher Al-Fe alloys in the range of 12–16 percent Al.

In 1902 Barrett, Brown, and Hadfield[1] investigated the magnetic properties of Al-Fe alloys up to 5.5 percent Al. They found that the resistivity increase for small additions of Al was more rapid than with any other element. Furthermore the magnetic properties compared favorably with the better Si-Fe alloys being produced at that time. Presumably the then greater cost of Al as compared to Si prevented a wider acceptance of these alloys containing low percentages of Al.

Yensen and Gatward[2] discovered Al did not materially increase the brittleness of Fe up to approximately 6 percent, whereas Si contents of 4.5 percent and above render the alloys quite brittle. Sykes and Bampfylde[3] reported being able to hot work Al-Fe alloys up to 17 percent Al. They also succeeded in drawing 12 percent Al-Fe wire but reported extreme difficulty in drawing a 14 percent Al-Fe alloy. During World War II, the Japanese substituted hot-rolled alloys containing about 16 percent Al for Ni-Fe alloys. Masumoto and Saito[4] report a maximum permeability of 55 000, an initial permeability of 3100, a coercive force of 0.04 oersted, and a remanence of 2100 gauss on hot-rolled

TABLE I. Typical chemical analysis of the alloying elements.

	C	Si	S	P	Mn	Ni	Cr	Cu
Electrolytic iron (typical)	0.0124	0.014	0.011	0.005	0.006	nil	nil	0.006
99.99% grade aluminum	nil	0.009	nil	nil	0.003	nil	nil	0.002
	V	Mo	Fe	Co	W	Al	O_2	Mg
Electrolytic iron (typical)	nil	0.002	99.92	nil	nil	nil	0.021	···
99.99% grade aluminum	nil	nil	0.005	···	nil	99.98	···	0.004

* The process and product, 16-Alfenol, described herein is a development of the U. S. Naval Ordnance Laboratory and is the property of the United States Government. Application for license under a pending U. S. patent application may be made to the Chief of Naval Research, Washington 25, D. C.

[1] Barrett, Brown, and Hadfield, Trans. Roy. Soc. Dublin 7, 67 (1902).

[2] T. D. Yensen and W. A. Gatward, University of Illinois Bulletin 95 (1917).

[3] C. Sykes and J. W. Bampfylde, J. Iron Steel Inst. (London) 130, 389 (1934).

[4] H. Masumoto and H. Saito, Sci. Rep. Research Inst. Tohoku Univ. 4, No. 4 (1952).

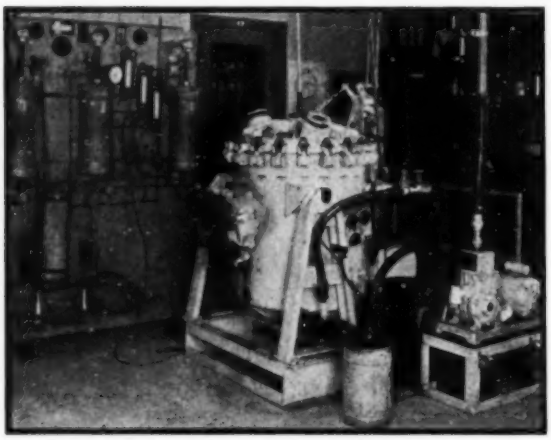


FIG. 1. Melting furnace and gas drying board.

16 percent Al-Fe alloys, approximately 0.020 in. thick. Attempts made in this country to produce similar electrical sheets by hot-rolling techniques resulted in low yields of rolled material.[5]

Despite reported earlier difficulties in processing, the 16 percent Al-Fe magnetic properties appeared so promising, that further investigation of this material was undertaken.

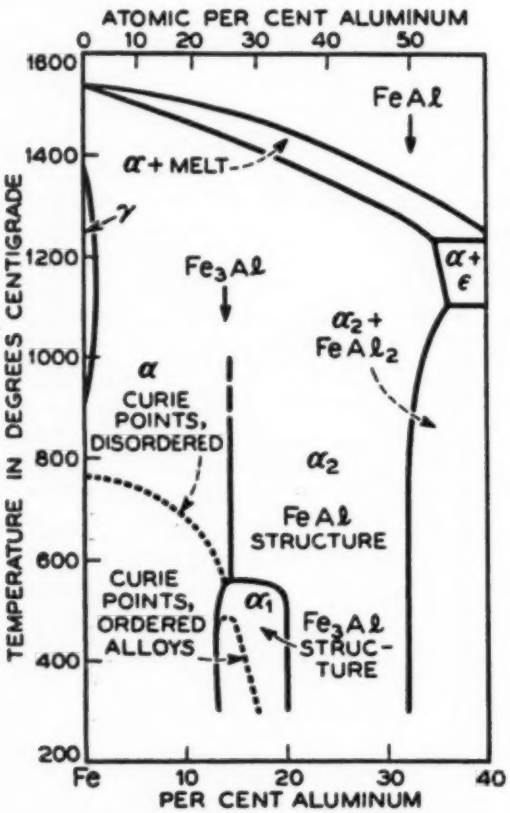


FIG. 2. Phase diagram of iron-aluminum alloys.

[5] G. H. Cole and M. F. Littmann, Armco Research Report (Armco-Navy Contract), July 30, 1952.

Preparation of Alloys

The alloy, which contains approximately 16 percent Al and 84 percent Fe, is prepared from electrolytic Fe and 99.99 percent grade ingot Al of the purity indicated in Table I. A controlled atmosphere furnace is used to insure a controlled composition.

The melting operation is carried out in a high-purity magnesium-oxide crucible and is accomplished by induction heating in a pressure-vacuum furnace of 30-pound capacity, shown in Fig. 1.

The furnace chamber is pumped down to a pressure of about 200μ prior to melting. This pressure will increase however as the Fe melts, due to the release of the dissolved gases from the melt. The molten Fe is given a decarburizing treatment with wet hydrogen, followed by treatment with dry hydrogen (dewpoint, -90°F or better) to effect deoxidation.

The hydrogen is purged from the chamber with pure dry helium, followed by evacuation to remove hydrogen dissolved in the melt during the decarburization and deoxidation treatments. The system is refilled with helium and the Al is added to the melt. The temperature is adjusted optically and the melt is chill-cast into slab molds coated with alundum cement mold wash. The mold cavity measures approximately 10 in.$\times$5 in.$\times$1 in. A typical chemical analysis of the cast alloy, including gaseous impurities, is given in Table II.

TABLE II. Typical analysis of hot-rolled 16 percent Al-Fe (0.125 in. thick).

Fe	Al	Si	S	P	Mn	C[a]	H_2	O_2	N_2
84.11	15.87	0.008	0.002	0.001	0.001	0.0025	0.00004[a]	0.00001	0.0002[a]

[a] Determined on ingot.

High-quality castings free from oxide inclusions and surface irregularities are necessary for the subsequent operations.

Chill casting is used to minimize segregation in the ingot. Segregation which does occur, however, is eliminated by hot working. This is shown in Table III.

Hot Rolling

The slabs are prepared for hot rolling by removal of the pipe or shrinkage cavity. Surface irregularities should not be removed by machining or grinding, because the machined or ground surface has a tendency towards transcrystalline cracking during subsequent hot rolling. These cracks are probably the result of stresses induced by the machining or grinding operations.

The entire slab is heated for one hour at 1000°C and rolled on a laboratory two-high mill. The slab is given four passes with reductions of approximately 0.005 in. for each pass. Five minutes are allowed for reheating between each set of four passes. This schedule is con-

tinued until a thickness of 0.375 in. is reached, after which only two passes of 0.005 in. each are given between reheating. The latter schedule is continued to a thickness of approximately 0.125 in. Rolling temperatures of 1100°C and above promote the formation of large grains.

Cold Rolling at Elevated Temperatures

The Al-Fe phase diagram of Fig. 2 shows that alloys in the range from approximately 10–20 percent by weight of Al have a tendency to order into an Fe_3Al type lattice (Fig. 3). At 13.9 percent (25 atomic percent) Al the entire alloy can exist as Fe_3Al. It has been reported that the Fe_3Al structure is mechanically softer than the disordered phase.[6] Based on this and other experimental evidence, it appeared advantageous initially to cold work this alloy in the order-disorder transformation region.

Ordering may be brought about by either a very slow cooling rate, on the order of 30°C/hr, from some annealing temperature above the critical order-disorder temperature range, or by heating for prolonged periods

TABLE III. Chemical composition of ingots before and after hot rolling.

Melt no.	Calculated composition	Bottom of slab	Middle of slab	Rolled material	
260	15.85% Al	16.26% Al	15.57% Al	15.89% Al	(0.014 in. thick)
259	15.85% Al	15.64% Al	...	15.87% Al	(After hot rolling to 0.125 in.)

of time in the region of 450°C to 560°C.[7] For the Fe_3Al composition (13.9 percent Al), the critical temperature is about 550°C.[7]

Hot-rolled material prepared as previously outlined was subjected to cold rolling at average temperatures of 500°C, 550°C, and 575°C. The latter temperature produced the most satisfactory results. Rolling at the lower temperatures of 500°C and 550°C produced edge cracking in the latter stages. The alloy was also more difficult to reduce at the lower temperatures. (It is possible that as a result of the cooling effect of the rolls, the alloy is being rolled at a temperature lower than indicated.)

One must be cautious in assuming that the order-disorder phenomenon is entirely responsible for the success attained in cold rolling this material. It is well known that in some metals at higher temperatures more slip planes become operative, allowing brittle materials to be worked. However, since the lower temperature limit of ductility seems to fall in the vicinity of the critical order-disorder temperature of this alloy, it is reasonable to assume that this phenomenon is at least

[6] A. J. Bradley and A. H. Jay, J. Iron Steel Inst. (London) **125**, 339 (1932).
[7] W. D. Bennett, J. Iron Steel Inst. (London) **171**, 372 (1952).

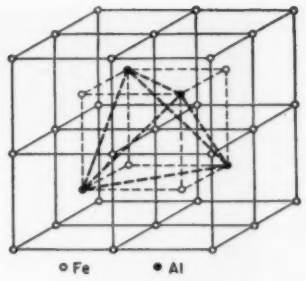


FIG. 3. Crystal structure of Fe_3Al.

partly responsible for the ability to cold work it at the observed temperature.

The rolling process, employing a temperature of 575°C, is initially carried out on a two-high mill allowing the alloy to reheat for approximately five minutes between passes. At a thickness of 0.040 in. the rolling is transferred to a four-high mill. At this point, the sheet may be sheared if a narrower strip is desired. The finish rolling is done on the four-high mill. Because of increased length the strip is heated in a long horizontal strip furnace, Fig. 4. The use of this type furnace also provides more efficient heating which expedites the rolling (2 minutes per pass). A back-tension reel is used on material under 0.014 in. to aid in maintaining straight strips. Using these techniques, it is possible to roll smooth, straight strips as thin as 0.004 in.

Photomicrographs (Figs. 5 and 6) of strip worked in the above manner definitely show a fibered structure which is typical of drastically cold-rolled metals and alloys. This highly fibered structure improves the toughness and ductility of the strip. Recrystallizing a piece of 0.007-in. material at 980°C for 5 minutes (Fig. 7) produced a more brittle material than a similar cold-worked piece. Material rolled at 575°C has a minimum surface oxidation.

It is possible to cold roll heavy gauges of this material at room temperature, provided it is sufficiently fibered to begin with. This method produces an extremely brittle strip, and is not recommended for that reason.

Thin-Gauge Rolling at Room Temperature

Strip material 0.007 in. thick, produced by the rolling process as outlined in the section on Cold Rolling at Elevated Temperatures, can be rolled at room temperature into very thin gauges. The 0.007-in. strip is annealed for a short time in a strip furnace and slit into suitable

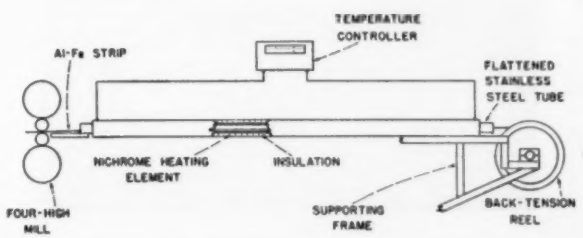


FIG. 4. Horizontal strip furnace.

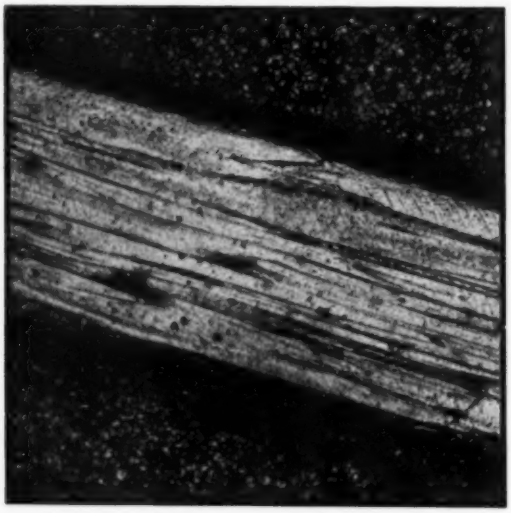


FIG. 5. 16 percent Al-Fe, cold rolled to a thickness of 0.028 in., magnification = 50×.

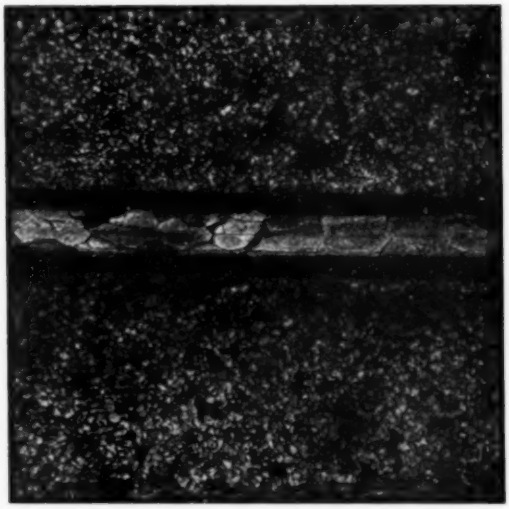


FIG. 7. 16 percent Al-Fe, cold rolled to a thickness of 0.008 in. and recrystallized (980°C for 5 min), magnification = 50×.

widths. The oxide coating formed during the previous treatment is removed electrolytically to minimize roll wear. The strip is then rolled on a cold-strip mill such as a Rohn mill, to a thickness of about 0.002 in. At this point any roughness on the edge is removed by slitting, and the rolling continued until the final desired thickness is attained. Figure 8 shows a 0.0005-in. strip rolled in this manner. Tape in thicknesses of 0.001 in. or below may be bent 180° upon itself and creased several times without breaking.

It has been observed that subjecting the 0.007-in. tape to heavy reductions in the first few passes in some cases improves the manner in which it rolls. This is probably the result of a heavy pass which brings the material down through the brittle stage very rapidly. Below 0.003 in. the material acquires considerable ductility, so that once this thickness is reached there is little difficulty in reducing it to the final desired thickness.

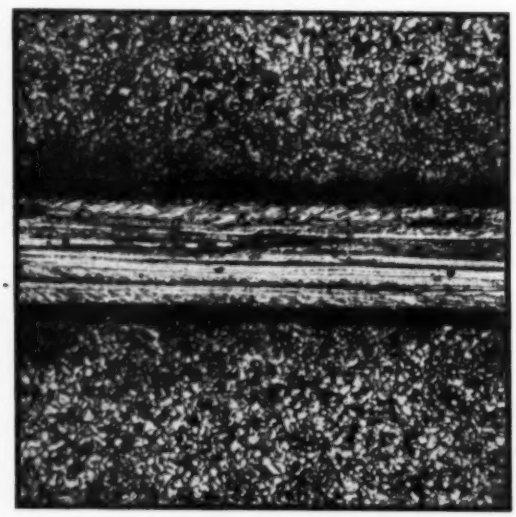


FIG. 6. 16 percent Al-Fe, cold rolled to a thickness of 0.013 in., magnification = 50×.

Static Magnetic Properties

Magnetic data which are presented here were obtained on a limited number of cores and heat treatments, however the information given shows the promising results which have thus far been obtained.

All the magnetic measurements were made on heat-treated laminations 2-in. o.d.×$1\frac{5}{8}$-in. i.d., punched from either 0.014-in. or 0.007-in. thick strips. Some laminations are shown in Fig. 8 along with a roll of strip material from which they were punched. Laminations, disks, or other desired shapes were punched with relative ease.

Table IV summarizes the magnetic results that were obtained after various heat treatments. All of these

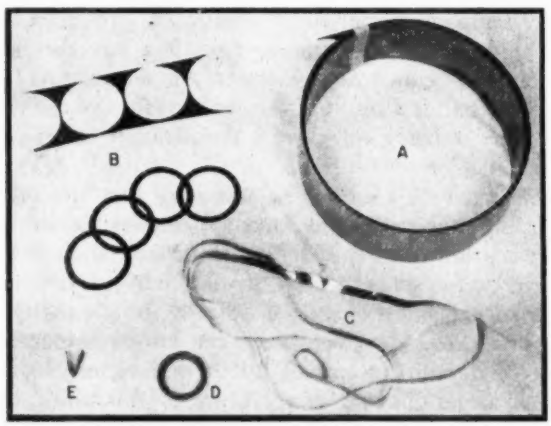


FIG. 8. Various forms of fabricated 16 percent Al-Fe. A, 0.014-in. thick rolled strip; B, stamped laminations; C, 0.0005-in. thick rolled tape; D, 0.001-in. thick wound core; E, 0.007-in. thick strip given sharp V bend.

TABLE IV. Magnetic properties of cold-rolled 16 percent Al-Fe alloys after various heat treatments.[a]

Specimen no.	Thickness of laminations	dc magnetic properties μ_{30}	μ_m	H_c	B_r	B_m[b]	Heat treatment
1992	0.014 in.	1745	15 800	0.104	3600	...	FH to 1000°C—hold 2 hr FC to 600°C—hold 10 min. RC
1992X[a]	0.014 in.	2778	115 000	0.024	4200	7600	FH to 900°C—hold ½ hr FC to 600°C—hold 10 min. Water quench
2018	0.014 in.	3056	67 700	0.036	4000	7400	FH to 1000°C—hold 2 hr FC to 600°C—hold 10 min. Water quench
2018X	0.014 in.	3418	36 500	0.049	3100	7800	FH to 1000°C—hold ½ hr FC to 600°C—hold 10 min. Oil quench
2044	0.007 in.	2868	74 900	...	4000	7900	FH to 900°C—hold 1 hr FC to 600°C—hold 10 min. Water quench
2044X	0.007 in.	3680	95 400	0.038	4100	7700	FH to 1000°C hold 2 hr FC to 600°C—hold 10 min. Water quench
2045	0.007 in.	4127	86 300	0.040	4000	7700	FH to 900°C—hold 1 hr FC to 600°C—hold 10 min. Water quench
2045X	0.007 in.	4158	77 800	0.044	4000	8000	FH to 1000°C—hold 2 hr FC to 600°C—hold 10 min. Water quench

[a] X—Signifies additional heat treatments on the original specimen. FH—Heated with the furnace. FC—Cooled with the furnace. RC—Withdrawn rapidly from heat zone to front of furnace.
[b] From $H_m = 30$ oersteds.

heat treatments were carried out in an atmosphere of pure dry hydrogen. During the annealing cycle the laminations were tightly clamped between two plates of high-purity iron. No insulation is necessary, inasmuch as the as-rolled 16 percent Al-Fe automatically forms its own tight insulating coating, composed almost entirely of Al_2O_3. This coating is not reduced by the hydrogen atmosphere even at the higher annealing temperatures. (Prior to quenching which follows the annealing operation, the annealing box is purged with helium to avoid the danger of explosion.)

Analysis of the data presented in Table IV, shows the best magnetic properties are obtained after a water quench, which means the alloy is in the disordered state. Initial permeabilities in the vicinity of 4000, coercive forces as low as 0.024 oersted, and maximum permea-

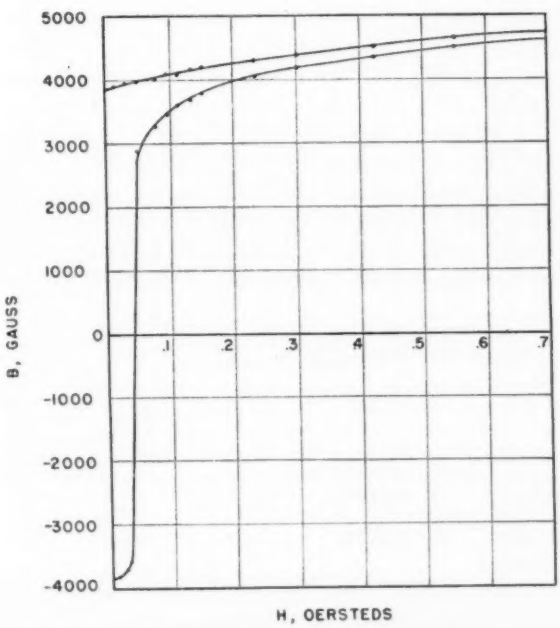


FIG. 9. Typical hysteresis loop (sample 2044X[2]); $\mu_{30} = 4880$, $\mu_m = 92\,800$, $H_m = 1.52$ oersteds, $B_m = 5070$ gauss, Loop area = 960 gauss-oersteds, Core loss = 76.4 ergs/cm³/cycle.

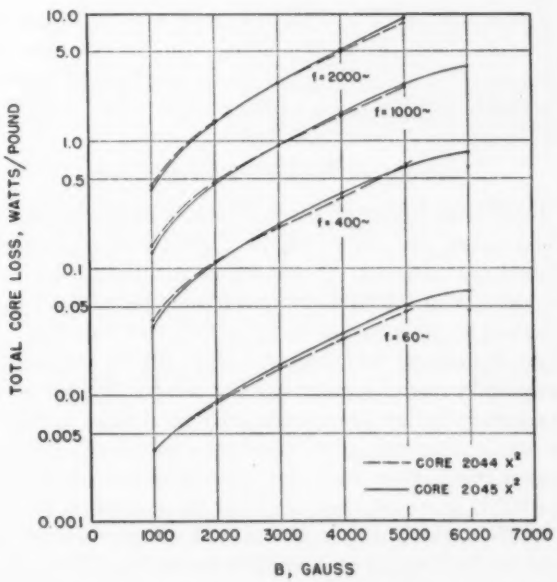


FIG. 10. Total core loss for 0.007-in. thick laminated cores of 16 percent Al-Fe.

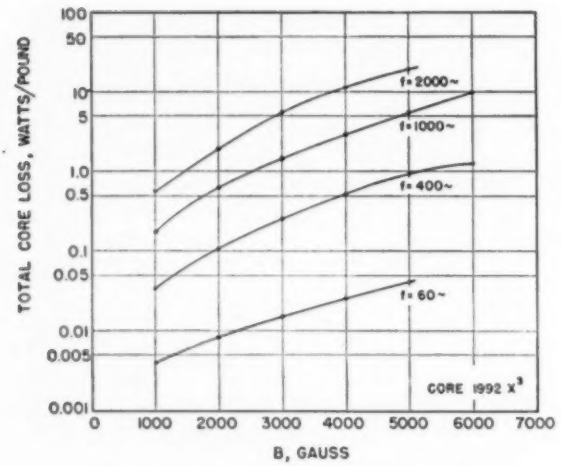


FIG. 11. Total core loss for 0.014-in. thick laminated cores of 16 percent Al-Fe.

bilities of over 100 000 have been obtained. The maximum permeability after the water quench is more than seven times the value obtained after the gas quench (rapid withdrawal from the heat zone to the front of the annealing box). It is evident that the rate of cooling through the critical order-disorder temperature is of importance.

Of interest is the extremely low anisotropy of the 16 percent Al-Fe alloy rolled at 575°C. A maximum value of about 1500 dyne-cm/cm³ was measured on a torque magnetometer[8] for this alloy heat treated for best magnetic properties. This compares to maximum values of 15 000–20 000 dyne-cm/cm³ obtained in magnetic torque measurements of hot-rolled Si-Fe alloys which are normally considered isotropic. Therefore this alloy shows promise for applications requiring nondirectional magnetic properties.

Figure 9 shows a typical hysteresis loop of 16 percent Al-Fe. Despite the low anisotropy of this material the *BH* loop is very rectangular, an advantage in applications such as magnetic amplifier cores.

Dynamic Magnetic Properties

Total core losses of several heat treated 16 percent Al-Fe cores, are presented in Figs. 10 and 11, which express the variation of core losses at different levels of flux density and frequency. No distortion of the loop is observed at frequencies up to 2000 cycles (see Fig. 12). These data show that it should be possible to substitute laminated cores of greater thicknesses for thinner gauge tape cores of other lower-resistivity materials, whenever the static magnetic properties are equivalent. Table V gives a comparison between grain-oriented silicon-iron tape 0.001 in. thick, and a typical 16 percent Al-Fe core of 0.007-in. thick ring lamination. A comparison of core losses at 5000 gauss on a volume basis (e.g., for a given core size) shows 16 percent Al-Fe alloy to be even more favorable because of its low density of approximately 6.5 g/cm³.

[8] A. M. Syeles and E. Adams, U. S. Naval Ordnance Laboratory Report No. 2152, September, 1951.

Other Physical Properties

This alloy has many important physical properties other than magnetic which heretofore have not been fully utilized because of the difficulty of fabricating the material into thin ductile sheets.

Its relative corrosion resistance in concentrated nitric-acid solutions was compared to that of 18–8 stainless steel. No perceptible action was noted when the 16 percent Al-Fe was immersed in the acid. The relative corrosion rates of the alloy and the stainless steel, on the basis of a test carried out for a total of 169 hours, are as follows:

	Loss in weight
16 percent Al-Fe	0.26 g/meter²/hr
18 8 stainless steel	0.0037 g/meter²/hr

On the basis of data taken from the Metals Handbook, 1948 edition,[9] 16 percent Al-Fe would be placed in category *B*, or the satisfactory resistant class of materials, for nitric acid.

Sykes and Bampfylde[3] carried out salt-spray tests for 800 hours with synthetic sea water, and found slightly less than twice the weight loss for an alloy with 16.16 percent aluminum as compared to that of a stainless steel.†

During hot rolling it was noted that the 16 percent Al-Fe could be heated to temperatures as high as 1200°C without noticeable scaling. An aluminum oxide coating is formed which protects the alloy from further oxidation. The average losses in weight of alloys containing 11 to 15 percent Al have been determined by oxidation tests at temperatures of 1000°C to 1350°C. The results listed in Table VII below are taken from Sykes and Bampfylde.[3]

The excellent resistance to oxidation of this alloy indicates its probable usefulness for applications at high temperatures, such as resistance heating elements and combustion chambers. Qualitative tests of heating elements made of thin ribbons from 16 percent Al-Fe show a comparatively long life for elements of this type.

TABLE V. Comparison of magnetic properties at varying frequencies.

Type core	Thickness	Core loss in *W*/lb at varying frequencies 60 cycles	400 cycles	1000 cycles
Oriented[a] Si-Fe tape	0.001 in.	···	1.2	3.2
16% Al-Fe laminations	0.007 in.	0.05	0.7	2.7

[a] Commercial catalog data. $B = 5000$ Gauss

[9] *Metals Handbook* (American Society for Metals, Cleveland, 1948), p. 561.

† Composition not specified.

The reported very low temperature coefficient of thermal conductivity[3] of a 16 percent Al-Fe alloy has been observed.

A relatively high resistivity of 153 micro ohm/cm has been measured on a strip of 0.014-in. 16 percent Al-Fe, heat treated for best magnetic properties. For alloys containing 11 percent Al and over, the change in resistivity with temperature over 600°C is very small.[3]

SUMMARY

Variations of hot-rolling procedures of 16 percent Al-Fe alloys were studied. The most successful method entailed the use of a hot-rolling temperature of 1000°C and about 0.005-in. reduction per pass. Temperatures higher than 1000°C produced an excessively large grain size that appeared to cause difficulties during cold rolling.

Cold rolling, at temperatures between 500°C and 700°C, was applied to the 0.125-in. thick hot-rolled material. After carefully observing the various cold-rolling characteristics at the temperatures above, it appeared that 575°C was the optimum temperature. This temperature is only slightly above the order-disorder temperature range and the success in cold rolling may be in part attributed to the ordering of the Al-Fe solid solution.

Thin-gauge cold rolling to a thickness of 0.0003 in was accomplished on 16 percent Al-Fe. This was done at room temperature by starting with 0.007-in. material and making rather heavy passes at the beginning of rolling. Highly polished thin-gauge tape was produced by this technique.

High yields of usable rolled stock were obtained by the fabrication techniques and melting procedures devised. The high permeability, rectangularity of the hysteresis loop, high resistivity, and the extremely low anisotropy obtained on 16 percent Al-Fe prove its usefulness as a magnetic material. For other applications it possesses such desirable properties as physical strength, high temperature oxidation resistance, high electrical resistivity, low thermal conductivity, and corrosion resistance to certain chemical solutions. In addition, 16 percent Al-Fe is a nonstrategic alloy possessing no elements that are in short supply.

TABLE VII. High-temperature oxidation rates.

Temperature	Loss in weight
1350°C	0.022 g/cm²/100 hr
1300°C	0.016 g/cm²/100 hr
1250°C	0.0070 g/cm²/100 hr
1200°C	0.0050 g/cm²/100 hr
1100°C	0.0020 g/cm²/100 hr
1000°C	0.0015 g/cm²/100 hr

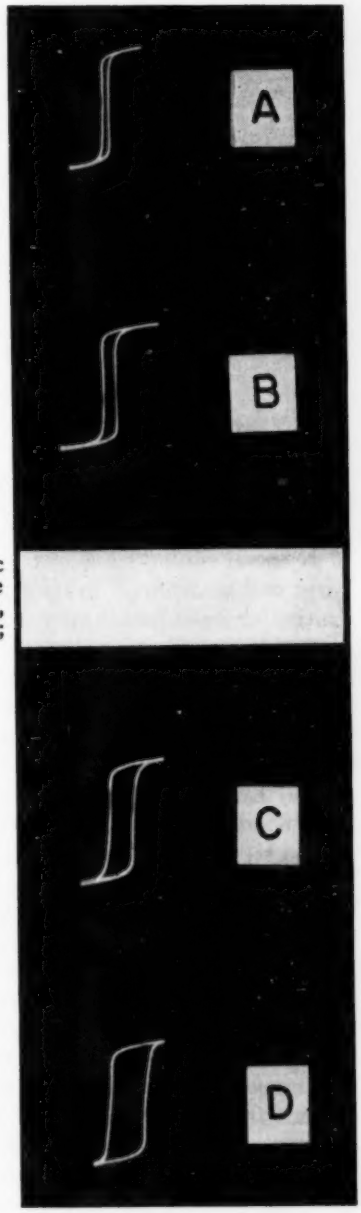


FIG. 12. Typical dynamic loops at 5000 gauss (sample 2044X[3]). A, 60-cycle; B, 400-cycle; C, 1000-cycle; D, 2000-cycle.

ACKNOWLEDGMENTS

The authors wish to extend their appreciation to Dr. G. W. Elmen for calling attention to the problem and for his helpful suggestions, and to members of the Magnetics Division, in particular to Mr. H. H. Helms and the members of the mechanical engineering group, for their services. These services were an important contribution to the success of this investigation.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 3 MARCH, 1954

Diode Theory in the Light of Hole Injection

JOHN A. SWANSON
International Business Machines Corporation, Poughkeepsie, New York
(Received June 25, 1953)

Classical diode theory, as applied to metal point contacts made to an n-type semiconductor, is adequate only for small forward voltages at which, under certain conditions, hole current may be negligible. The shape of the theoretically predicted diode characteristic is not affected by the hole injection process at low voltages, no matter what the composition of the current. However, for voltages in excess of a certain value, dependent only on the resistivity of the material (on the order of 0.1 volt for 5 ohm-centimeter germanium), the spreading resistance is comparable to the barrier resistance, and it is the hole injection process which accounts for continued rectification. The extent to which the spreading resistance is decreased by hole injection depends on the ratio, γ, of hole current to total current. The present paper includes a theory of the effect of this ratio on the diode characteristic at higher forward voltages. A method of measuring γ from V–I characteristics alone is indicated.

INTRODUCTION

ALTHOUGH the phenomenon of hole injection by metal contacts made to n-type semiconductors is fairly well understood in its broader aspects, no quantitative or semiquantitative discussion of the static diode characteristic, as modified by hole injection, seems to exist in the literature. J. Bardeen and W. H. Brattain[1] have given a valuable discussion of the matter, but their treatment is restricted for simplicity to the special case in which all of the forward current is carried by holes. The present paper treats the more general problem in which the behavior of the fractional hole current is to be deduced on theoretical grounds; however, the descriptive equations are complex, and for this reason the discussion is often semiquantitative.

To prepare the way for the subsequent analysis, a short summary of diode theory[2] will be given here, in the course of which some essential formulas used in the later sections will be obtained. According to the most widely accepted model of the contact between a metal and an n-type semiconductor, an electrostatic barrier extends from the metal surface into the semiconductor a distance less than the mean free path, but many times the length of a unit cell of the crystal. The barrier to electrons incident either from the metal or the semiconductor exists by virtue of the forbidden energy region in the semiconductor. Although the field immediately inside the semiconductor may be quite large, analysis indicates that for barriers of the height usually obtained in practice there should be no significant quantum-mechanical penetration in the forward direction. Electrons must then have an energy equal to the barrier height in order to traverse it. Holes formed in the semiconductor next to the surface of the metal also encounter a barrier unless the energy corresponding to the vacant state is less than the energy of the top of the valence band at its lowest point.

In order to obtain notational simplicity and symmetry, we shall define the height of a barrier as the absolute value of the potential difference between the barrier extremity and the Fermi level in the metal. The height of the electronic barrier will be represented by the symbol "ϕ_b," and that of the barrier to holes by "ϕ_v," as indicated in Fig. 1. The symbol "ϕ_c," which also occurs in Fig. 1, denotes the height of the bottom of the conduction band above the Fermi level. It will prove convenient to define the quantity $(\phi_b-\phi_c)$ as the "barrier to semiconductor electrons"; the meaning of the term "barrier" will thus occasionally be modified by the context.

Provided that the Fermi level lies in the energy gap at a distance large compared with thermal energy from the edge of each allowed band, holes and electrons both obey Boltzmann statistics. The currents across the interface between metal and semiconductor at equilibrium consequently have a simple exponential dependence on the barrier height:

$$I_{0n}=(1-r_n)SA_nT^2\exp(-\alpha\phi_b), \quad (1)$$

$$I_{0p}=(1-r_p)SA_pT^2\exp(-\alpha\phi_v). \quad (2)$$

Equations (1) and (2) express the currents of electrons (indicated by the subscript n) and of holes (indicated by the subscript p) which flow into the metal from the semiconductor at equilibrium due to the random motion of the particles. Equal and opposite currents must flow from the metal at equilibrium in order to satisfy the requirement of detailed balancing. The current flowing from either side of the interface is given the symbol I_0, with appropriate subscripts. The quantity r is the average reflection coefficient for particles incident from the semiconductor, S is the area of the contact, and T is the absolute temperature. Each A

[1] J. Bardeen and W. H. Brattain, Phys. Rev. 75, 1208 (1949). The experimental evidence at that time also favored unit hole injection efficiency, but this evidence was later found to be incorrectly interpreted due to an uncertainty in the carrier mobility ratio (see reference of footnote 16). That the injection efficiency cannot be unity is shown by the failure of the diode exponential law at comparatively low voltages, whereas for unit injection efficiency the diode characteristic should continue to have an exponential form at higher voltages (see section IV).

[2] For a review of the development of this subject, including reference to the pioneer work of Mott and Schottky, see H. C. Torrey and C. A. Whitmer, *Crystal Rectifiers* (McGraw-Hill Book Company, Inc., New York, 1948), first edition, Chap. 4.

is defined as equal to the thermionic emission constant (120 amp/deg² cm²) times the ratio of the appropriate effective mass to the electronic mass. The quantity α is by definition equal to q/kT, where q is the absolute value of electronic charge, and k is Boltzmann's constant. A derivation of Eq. (1) can be found in most books dealing either with diode theory or with thermionic emission,[3] and (2) is a result following from the symmetry of holes and electrons.

When a bias V_b is applied to the barrier (by convention V_b is positive when the metal terminal is positive), the electronic barrier ϕ_b is assumed to remain fixed. If there are no semiconductor surface states, the electric field is continuous across the surface of the semiconductor; a change in this electric field results in a much larger change in the barrier potential than in the potential difference between metal and semiconductor, since the distance separating metal and semiconductor is of the order of atomic dimensions, while the barrier itself extends over many atoms. If surface states do exist, they should further stabilize ϕ_b provided that the occupation of surface states is controlled by the Fermi level in the metal.[4] It is clear that, granted the condition in question, the barrier height for electrons leaving the semiconductor is decreased by the amount of the bias V_b (Fig. 2). If the electron concentration just beyond the barrier is assumed constant (later it will be necessary to take into account the increased electron concentration necessary to balance the positive charge of the injected holes), the current given by Eq. (1) is increased by the factor $e^{\alpha V_b}$. The flow of electrons from metal to semiconductor remains constant, so that the net electronic current I_n, oppositely directed from the average motion of the electrons, is given by

$$I_n = I_{0n}(e^{\alpha V_b} - 1). \tag{3}$$

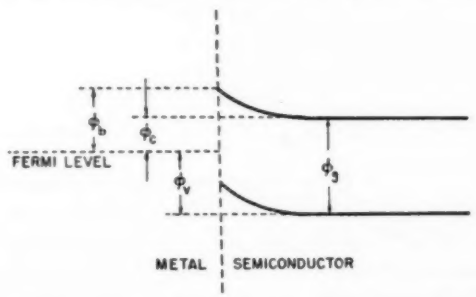


FIG. 1. Disposition of semiconductor energy bands at a metal-semiconductor interface at equilibrium. The heavy lines represent the edges of the valence and conduction bands, which are separated by the energy $q\phi_g$.

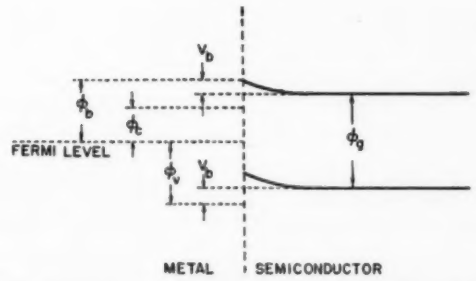


FIG. 2. Disposition of semiconductor energy bands upon the application of a positive voltage to the metal. A part of this voltage, V_b, falls across the barrier.

Until recently the theory neglected the hole current, so that Eq. (3) was expected to govern the diode characteristic. The total voltage, V, is composed of two parts: that falling across the barrier, or V_b, and that falling across the rest of the semiconductor, or V_s. The resistance of the bulk of the semiconductor, called the "spreading resistance" because of the current distribution near the contact, will be designated by R_s. Then the expected form of the diode characteristic is:

$$I = I_{0n}(e^{\alpha(V - IR_s)} - 1), \tag{4}$$

where I is the total current in the one carrier theory.

In the light of present knowledge formula (4) is inadequate, since it does not include the contribution arising from the hole injection process. It is also inadequate from the point of view of experiment: one finds a qualitatively similar dependence of current on voltage, but this dependence is not quite exponential; in order to fit the experimental curve with an expression of the form (4), it is necessary to allow α to be a variable which decreases as the voltage increases.[5,6] This decrease in the "effective alpha" has been ascribed to nonuniform electronic barriers; i.e., to barriers having different heights ϕ_b at different points of the contact. At those points having low barriers the barrier to semiconductor electrons ($\phi_b - \phi_c$) may be completely nullified by the applied voltage at comparatively low voltages; the current through these regions (assumed small in area) then becomes saturated, and does not contribute to the rectification at higher voltages. This explanation of the failure of the diode current to follow a strictly exponential law, which was proposed by Bethe and developed by Johnson, Smith, and Yearian,[7,8]

[3] See, for example, F. Seitz, *The Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940), first edition, p. 164. It may be of interest to point out that the semiconductor plays the role of the vacuum in the analogy between diode theory and thermionic emission, and that the correction factor m^*/m follows from the presence of a linear factor in m in the thermionic emission constant. The over-all reflection coefficient r compensates for the total reflection of some electron waves, as well as for ordinary reflections, and there is consequently no relation of equality between the over-all reflection coefficients on each side of a crystal interface. The total reflections arise from the fact that an electron state having specified energy and specified tangential components of crystal momentum may sometimes exist on one side of the boundary, while failing to exist on the other.

[4] For a qualitative discussion of the effect of surface states, see W. Shockley, *Electrons and Holes in Semiconductors* (D. Van Nostrand Company, Inc., New York, 1950), first edition, p. 96. Surface states as an explanation of anomalous surface properties of semiconductors were proposed by Bardeen (Bardeen, Phys. Rev. **71**, 717 (1947)).

[5] H. J. Yearian, J. Appl. Phys. **21**, 214 (1950).

[6] Torrey and Whitmer, *Crystal Rectifiers* (McGraw-Hill Book Company, Inc., New York, 1948), first edition, Sec. 4.4.

[7] H. A. Bethe, "Theory of the boundary layer of crystal rectifiers," RL Report No. 43-12, November 28, 1942.

[8] Johnson, Smith, and Yearian, J. Appl. Phys. **21**, 283 (1950).

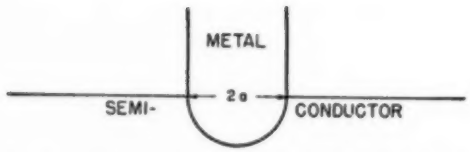


FIG. 3. Contact geometry, as idealized for simplicity. The hemispherical metal point is assumed embedded in the semiconductor, which has a plane surface, and is infinite in extent.

has been called the "multicontact theory." Unfortunately, in the absence of knowledge concerning the actual variation of ϕ_b, it does not afford a quantitative prediction of the diode characteristic. It will later be shown (Section IV) that the effect of hole injection is such as to fix a minimum value of the ratio of spreading resistance to barrier resistance, and consequently a maximum voltage for which the "classical" diode characteristic (i.e., the exponential form (4) without the spreading resistance term) should be valid. This effect probably outweighs the influence of multicontacts for the high purity materials typically in use at the present time. Consequently, a uniform barrier will be assumed in carrying through the analysis.

An obvious conclusion to be drawn from Eq. (4) is that for sufficiently high voltages the spreading resistance controls the current. Since the injection of holes tends to decrease the spreading resistance because of the extra conductivity provided by the additional carriers, at least one effect of hole injection will be the prolongation of the rectifying action of the diode, and a diminution in the final limiting resistance of the diode. These effects were observed before the existence of hole injection was recognized.[9,10]

In order to predict the quantitative nature of the high current effects, and to understand the role played by holes at moderate currents, it is necessary to develop a theory relating the hole current to the voltage across the barrier. This could be done on the basis of Eq. (2), taking into account the change in concentration of holes next to the emitter; in that case this change must be regarded as a temporarily unknown quantity, to be determined later by the equations relating hole current and hole concentration in the spreading resistance region. Although this procedure is correct and more general than the one to be followed, it is simpler to lay down the restriction that the net hole current be much smaller than the total current, $I_{0p}\exp(\alpha V_b)$, which passes from the metal into the semiconductor. This restriction will usually be obeyed, since the diffusion of holes into the interior of the semiconductor is a slower process than their creation at the interface. Accordingly, the holes next to the metal are very nearly in equilibrium with the unfilled states in the metal, so that the distribution of holes near the contact is appropriate to the Fermi level in the metal. As long as the energy difference between this Fermi level and the top of the valence band is several times kT, holes will obey Boltzmann statistics; we conclude that

$$p = p_0 \exp(\alpha V_b), \tag{5}$$

and hence that

$$\Delta p = p_0(\exp(\alpha V_b) - 1), \tag{6}$$

where p_0 is the equilibrium concentration, and Δp is the change in concentration just beyond the barrier when a voltage V_b occurs across the barrier.

We reserve the treatment of large hole currents for the later sections. If the applied voltage is small, however, the hole current will be principally due to diffusion, since the electric field will be small. If the contact is assumed to have a hemispherical shape (Fig. 3), the diffusion equation is easily integrated, giving:

$$I_p = \frac{2\pi a^2 q\mu}{\alpha}\left(\frac{1}{a}+\frac{1}{L}\right)\Delta p. \tag{7}$$

Here μ is the mobility of holes; a, the radius of the contact; and L, the diffusion length (L is equal to $(\mu\tau/\alpha)^{\frac{1}{2}}$, where τ is the lifetime of excess holes). The differential equation leading to (7) is a special case of Eq. (10). In view of Eqs. (3) and (6), the total current may be written as:

$$I = (I_{0n}+I_{ep})(\exp(\alpha V_b)-1) \tag{8}$$

$$\doteq I_0(\exp(\alpha V)-1), \tag{8'}$$

where I_{ep} is defined as:

$$I_{ep} \equiv 2\pi a^2 q\mu\alpha^{-1} p_0(1/a+1/L) \tag{9}$$

$$\doteq 2\pi a q\mu p_0/\alpha, \tag{9'}$$

and

$$I_0 \equiv I_{0n}+I_{ep}.$$

The variation of the hole current with barrier voltage has exactly the same form as the variation of the electron current at low currents, although the constant multiplying the exponential factor is not I_{0p}, the equilibrium current flowing across the interface from both sides, but a smaller effective constant I_{ep} determined solely by the geometry of the contact and the properties of the bulk semiconductor. Equation (8′) is an approximate form of (8), valid if the voltage across the spreading resistance is assumed negligible. With this restriction the usual theoretical diode characteristic does not have to be modified when the hole injection process is taken into account. In section IV we shall discuss the upper voltage limit to the validity of Eq. (8′).

Equation (9′), an approximate form of (9), is valid only if the diffusion length is much larger than the radius of the contact. Practically, this means that for typical contacts (10^{-3} cm$<a<10^{-2}$ cm) the lifetime of excess carriers must be larger than a tenth of a microsecond. Such lifetimes are almost always to be expected; furthermore, the assumption of lifetimes of this

[9] Bray, Lark-Horovitz, and Smith, Phys. Rev. **72**, 530 (1947).
[10] An explanation on the basis of hole injection was given by E. J. Ryder and W. Shockley, Phys. Rev. **75**, 310 (1949).

order of magnitude or greater allows a considerable simplification of the discussion. In fact, we may totally neglect recombination effects for low currents, as shown by the validity of (9′) for $L \gg a$, and we small assume that recombination effects may also be neglected at high currents. The equations governing hole and electron flow in the bulk germanium may then be solved for arbitrary currents, as discussed in the next section.

I. SPREADING RESISTANCE

We have derived an equation relating the electron current to the barrier voltage [Eq. (3)] and one relating the concentration of holes just beyond the barrier to the barrier voltage [Eq. (6)]. A complete solution can be obtained only if we find the dependence of the electron current, the hole current, and the hole concentration upon the voltage V_s across the spreading resistance. In principle, it will then be possible to eliminate all quantities that are not directly measurable from the system of equations, and thereby to derive a relationship between the total current and the total voltage across the diode.

The differential equations governing steady state current flow in a semiconductor of uniform impurity content and having infinite excess carrier lifetime are:

$$i_p = -\mu q[(p_0+p_1)\nabla\Psi+\alpha^{-1}\nabla p_1], \tag{10}$$

$$i_n = -b\mu q[(n_0+p_1)\nabla\Psi-\alpha^{-1}\nabla p_1], \tag{11}$$

$$\nabla\cdot i_p = 0, \tag{12}$$

$$\nabla\cdot i_n = 0. \tag{13}$$

In these equations i_p and i_n are the hole and electron current densities, respectively, Ψ is the electrostatic potential, p_1 is the change in each carrier concentration from the equilibrium values p_0 and n_0, and b is the ratio of electron mobility to hole mobility. The concentration changes are the same for each carrier, since space charge neutrality must be preserved in the interior of the semiconductor to a high degree of approximation. In both (10) and (11) the first terms represent the conduction current, and the last terms, the diffusion current. Equations (12) and (13) express the fact that holes and electrons are individually conserved. Arguments leading to this set of equations may be found in the book of Shockley.[11]

Although these equations are nonlinear, the simplified geometry we have assumed makes an exact solution possible. The method of solving is given in Appendix I, while the solution itself is exhibited below as a pair of simultaneous equations.

$$p_1 = n_b(e^{\alpha^*\Psi}-1), \tag{14}$$

$$p_1 = \frac{\alpha}{2}\left(n_0 b_- R_{sn} I\frac{a}{r} + n_-\Psi\right). \tag{15}$$

[11] W. Shockley, *Electrons and Holes in Semiconductors* (D. Van Nostrand Company, Inc., New York, 1950), first edition, Chap. 12.

The symbol r stands for the radial distance of a point in the semiconductor from the center of the hemispherical contact; i.e., $(r-a)$ is the distance from the contact surface itself. Both $p_1(r)$ and $\Psi(r)$ have been required to vanish at $r=\infty$. The quantity R_{sn} is the spreading resistance of the contact calculated as if the only current carriers were electrons. It would be approximately equal to the actual spreading resistance in sufficiently n-type samples, were there no hole injection. In the latter case R_{sn} is equal to $\rho/2\pi a$, where ρ is the resistivity of the material. The quantity I is the total current. The other parameters are either constants or functions of the composition of the current. Letting γ stand for that fraction of the current carried by holes, we define them as follows:

$$b_+ = b\gamma+(1-\gamma),$$

$$b_- = b\gamma-(1-\gamma),$$

$$n_+ = n_0+p_0 \doteq n_0,$$

$$n_- = n_0-p_0 \doteq n_0,$$

$$n_b = \frac{n_+b_-+n_-b_+}{2b_-} \doteq \frac{b\gamma}{b_-}n_0,$$

$$\alpha^* = \frac{b_-}{b_+}\alpha.$$

The approximate forms are suitable for highly n-type material. In terms of the constants of the material the electronic spreading resistance is

$$R_{sn} = \frac{1}{2\pi ab\mu q n_0}.$$

Because of the complex way in which the parameter γ enters Eqs. (14) and (15), the interpretative and manipulative problem presented by them is rather formidable. An initial simplification can be obtained, however, by utilizing the approximations valid for highly n-type material. Since in most physical cases of interest n_0 is more than a hundred times as large as p_0, these approximations should be very good. Noticing that our former definitions of Δp and V_s require that

$$p_1(a) = \Delta p,$$

$$\Psi(a) = V_s,$$

we obtain

$$\Delta p = \frac{b\gamma}{b_-}n_0(e^{\alpha^*V_s}-1), \tag{16}$$

$$\Delta p = \frac{\alpha n_0}{2}(b_-R_{sn}I+V_s). \tag{17}$$

Equations (16) and (17) can be combined to yield a formula not containing Δp,

$$I = \frac{1}{b_-R_{sn}}\left[\frac{2b\gamma}{b_-\alpha}(e^{\alpha^*V_s}-1)-V_s\right]. \tag{18}$$

Equations (16) and (17), combined with those derived in the introduction relating the electronic current and the hole concentration with the change in the barrier height upon the application of voltage, constitute a system of equations giving the diode characteristic in implicit form. In the succeeding sections we shall attempt to draw some general conclusions from this system of equations, but for the present we confine our attention to the spreading resistance alone.

The spreading resistance R_s is defined as the ratio V_s/I. Inspection of (18) reveals that R_s is, in general, not a constant, and that the behavior of R_s depends radically on the fractional hole current γ. We shall first examine the form (18) takes in the limit of low voltages. The exponential factor becomes linear in V_s, and we find:

$$V_s = b_+ R_{sn} I, \tag{19}$$

or

$$R_s = [1+(b-1)\gamma]R_{sn} \quad \text{(low voltages)}. \tag{20}$$

R_s increases linearly with the fractional hole current for low voltages, an odd result which, however, is not difficult to understand; for the electron diffusion current is larger than the hole current by the factor b and in the reverse direction, thus requiring a counteracting electric field larger by the factor b than the field required to produce an equivalent pure electron current.

The behavior of R_s for high voltages is perhaps of greater interest. We note that α^* is negative for $\gamma<1/(1+b)$, and positive for $\gamma>1/(1+b)$. The sign of α^* is determined by the sign of b_-, since b_+ is positive definite. The critical value $1/(1+b)$ will be denoted by the symbol "γ_c." Let ϵ be a positive quantity not zero, and let $(\gamma_c+\epsilon)$ be a lower limit for γ in a given case. It is clear that the exponential in (18) must eventually dominate, as the voltage across the spreading resistance becomes large without limit. The actual spreading resistance, V_s/I, is then approximated by

$$R_s \doteq \frac{b_-^2 R_{sn}}{2b\gamma} \frac{\alpha V_s}{e^{\alpha^* V_s}-1} \quad (\gamma>\gamma_c+\epsilon). \tag{21}$$

In other words, if the hole injection ratio γ remains always greater than γ_c by a definite, finite amount, the spreading resistance acts as a rectifier, the current depending exponentially upon the voltage. In fact, if γ is constant and greater than γ_c, the spreading resistance will act as an ideal rectifier for high voltages, except that the "effective alpha" occurring in the exponent is reduced from its ideal value, q/kT, by the factor b_-/b_+.

If γ is less than γ_c, and has an upper limit $(\gamma_c-\epsilon)$, where ϵ is again a positive, nonvanishing quantity, the spreading resistance has the value

$$R_s \doteq -b_- R_{sn} = [1-(b+1)\gamma]R_{sn} \quad (\gamma<\gamma_c-\epsilon) \tag{22}$$

for large V_s, since α^* is negative, and the exponential factor is insignificant. The spreading resistance accordingly has the lower limit $\epsilon(b+1)R_{sn}$ and will tend to approach a constant value if γ approaches a number less than γ_c.

If the quantity $|\gamma-\gamma_c|$ does not remain larger than a positive number ϵ, but, on the contrary, approaches zero, Eq. (22) leads us to suppose that R_s will also approach zero for high voltages. The special case for which γ approaches γ_c is actually of great interest, since whenever large currents flow there is a tendency for the current to be due principally to conduction rather than to diffusion, and hence for γ to be equal to γ_c. In order to treat this case, we must expand (18) in a Taylor's series, and apply the appropriate limiting procedures. The quantity b_- does not appear in the first two terms of this expansion, which can be written concisely as:

$$I = \frac{V_s}{b_+ R_{sn}}\left(1+\frac{b\gamma}{b_+}\alpha V_s\right) + \frac{2b\gamma}{b_-^2 \alpha R_{sn}}(e^{\alpha^* V_s})_{03}.$$

The subscript on the exponential in the second term signifies that terms of order lower than three in the Taylor's expansion are to be omitted since they have already been included as a part of the first term.

The second term vanishes in the limit of high voltage if the limit of $b_- V_s^3$ is zero. The reader may verify that if $b_- V_s$ approaches a finite limit, including zero, I will vary asymptotically as V_s^2; the asymptotic dependence will be of higher or lower order depending on whether $b_- V_s$ becomes positively or negatively infinite. For the special case $\gamma=\gamma_c$, we have:

$$I = \frac{1+b}{2b}\frac{V_s}{R_{sn}}\left(1+\frac{\alpha V_s}{2}\right) \quad (\gamma=\gamma_c). \tag{23}$$

The spreading resistance thus acts as a rectifier for $\gamma>1/(1+b)$, but remains finite and nonvanishing for $\gamma<1/(1+b)$. For $\gamma=1/(1+b)$, the resistance decreases inversely as the voltage across it for sufficiently high voltages.

To conclude this discussion of the spreading resistance we take note of a difficulty connected with the assumption of space charge neutrality in the interior of the semiconductor. The most natural boundary between the barrier and the spreading resistance region is the spherical surface on which Ψ has a maximum. However, the solutions we have obtained do not satisfy the boundary condition $E=0$ at $r=a$. This difficulty, often ignored in the literature, can be overcome only by dropping the condition $p_1=n_1$, and setting Poisson's equation in its place. The assumption of space charge neutrality is hence a poor approximation in the immediate vicinity of the barrier. However, the inclusion of Poisson's equation makes the mathematical problem quite formidable; also, since considerable deviation from space charge neutrality occurs in a small region only, we do not anticipate any serious loss of accuracy from this source.

II. HIGH CURRENTS

We shall understand the phrase "high currents" to imply a hole concentration in the neighborhood of the emitter much larger than the equilibrium electron concentration; however, the barrier for electrons leaving the semiconductor must retain a height of several times the average thermal energy of electrons in spite of the applied voltage. Only if the last condition holds is it permissible to use the undisturbed Boltzmann distribution to calculate the electron current from semiconductor to metal. It may be that in particular cases the hole concentration will never become considerably larger than the equilibrium concentration because the barrier is too low; then "high currents" in our terminology will not exist. The phrase "very high currents," or "saturation currents," will signify those currents which flow when the barrier to semiconductor electrons has been virtually removed by the applied voltage.

In the following, the qualitative behavior of γ as the current is increased from low to high values will be investigated. It will be found that γ approaches γ_e in such a way that the product b_-V_s approaches a finite limit. Naturally V_s cannot become infinite in practice; the theory we are employing is no longer applicable when the barrier has been completely nullified by the applied voltage; however, by assuming the product b_-V_s to be approximately equal to its limit for high currents it is possible to derive a formula for the diode characteristic which ought to be more nearly correct over its range of validity than formulas derived on the basis of a fixed γ hypothesis. In addition, the discussion shows that the fractional hole current will increase with current if its initial value is less than γ_e. This behavior of γ is experimentally observed.

When the injected hole density is comparable to the original electron density, Eq. (3) has to be modified as follows:

$$(1-\gamma)I=I_{0n}(e^{\alpha V_b}-1)+\frac{\Delta p}{n_0}I_{0n}e^{\alpha V_b}. \tag{24}$$

The extra term represents the contribution of the extra electrons next to the barrier necessary to preserve space charge neutrality.[12] The expression $(1-\gamma)I$ is equivalent to I_n. This equation determines the diode characteristic implicitly when taken in conjunction with Eqs. (16) and (17) and the trivial one:

$$V=V_b+V_s. \tag{25}$$

Remembering that $\Delta p=p_0(e^{\alpha V_b}-1)$, we see that (24) may be replaced by

$$(1-\gamma)I=I_{0n}\frac{\Delta p}{p_0}\left(1+\frac{\Delta p}{n_0}\right), \tag{26}$$

where a term p_0/n_0 has been omitted, since it is small compared to unity.

Equations (16) and (17) can be combined to yield a relation involving only Δp, γ, and I.

$$\frac{\Delta pb_-}{n_0b\gamma}-\frac{b_+}{2b\gamma}\log\left(1+\frac{\Delta pb_-}{n_0b\gamma}\right)=\frac{b_-^2}{2b\gamma}\alpha R_{sn}I. \tag{27}$$

Equation (27) shows that the added hole density Δp behaves in distinctly different ways as the current becomes large without limit, according to whether $\gamma>\gamma_e$, or $\gamma<\gamma_e$. In the first case, Δp must also become infinite; whereas in the second, Δp remains less than $-n_0b\gamma/b_-$, approaching that quantity in the limit. Since Δp can become much larger than n_0 provided that the barrier is high enough, we can conclude that b_- must approach zero if the low current γ, which we shall call γ_0, is less than γ_e. In other words, γ approaches γ_e if its initial value is less than γ_e, but the barrier is still high enough to allow large injected densities.

If $\gamma_0>\gamma_e$, the argument is not so simple. We first note that the condition implies a high barrier, and thus that $\Delta p\gg n_0$ for sufficiently high voltages. We accordingly contract (26) by omitting the linear term, and substitute the resultant expression for I into (27). We find that

$$1-\frac{b_+}{2b\gamma}\frac{1}{x}\log(1+x)=\frac{\alpha R_{sn}I_{0n}n_0}{2p_0}\frac{b\gamma}{1-\gamma}x, \tag{28}$$

where x represents the quantity $b_-\Delta p/n_0b\gamma$.

If b_- remains positive and nonvanishing, x must approach infinity as Δp approaches infinity. However, x cannot approach infinity, for the left side of (28) is bounded for $\gamma>\gamma_e$, while the right is not. It follows that b_- must vanish in the limit of high Δp, or that γ approaches γ_e for this case also. We conclude that x approaches a limit

$$\lim_{\Delta p\to\infty}x=\lambda \quad (\lambda>-1), \tag{29}$$

and that

$$\lim_{\Delta p\to\infty}\gamma=\gamma_e. \tag{30}$$

Both $b_+/2b\gamma$ and $b\gamma/1-\gamma$ are unity when $\gamma=\gamma_e$. Therefore, λ may be found from the equation:

$$1-\frac{1}{\lambda}\log(1+\lambda)=\frac{\phi}{2}\lambda, \tag{31}$$

where $\phi\equiv\alpha R_{sn}I_{0n}n_0/p_0$, which is simply $1/b$ times the ratio of electronic current to hole current for low voltages. If $\phi=1$, $\gamma_0=\gamma_e$, and $\lambda=0$. It is not difficult to find an exact solution for this case; γ remains equal to γ_e if its initial value is γ_e (see Appendix II).

[12] As discussed previously, space charge neutrality is not approximated as well near the barrier as it is throughout most of the spreading resistance region. The present term is thus perhaps an overcorrection.

Multiplying (26) by $(b_-)^2/b\gamma$, and taking the limit we find

$$\lim_{\Delta p\to\infty}\left(\frac{b_-^2 I}{b\gamma}\right)=\frac{n_0 I_{0n}}{p_0}\lambda^2. \tag{32}$$

Similarly, multiplying (17) by $b_-/b\gamma n_0$ and taking the limit we have, using (32),

$$\lim_{\Delta p\to\infty}(b_-\alpha V_s/b\gamma)=2\lambda-\phi\lambda^2$$
$$=2\log(1+\lambda). \tag{33}$$

To get an asymptotic approximation for the high current diode characteristic we combine (32) and (33), ignoring the limit sign, and obtain

$$I=\frac{(b+1)\alpha\phi}{4bR_{sn}}\left(\frac{\lambda}{\log(1+\lambda)}\right)^2 V^2, \tag{34}$$

in which we have replaced V_s by V. This is allowable since, as shown in section IV, most of the voltage falls across the spreading resistance when the injected concentration is much larger than the normal majority concentration.

We next consider briefly what to expect if the deviation from charge neutrality near the barrier requires us to place a smaller exponent than unity on the second of the Δp's occurring in (26). We should then have

$$(1-\gamma)I\doteq k(\Delta p)^\beta \tag{35}$$

at high currents, with β a number less than 2, and k an appropriate constant. An analysis similar to that carried out above shows that in this case $b_-^{1/\beta-1}\Delta p$ has a finite positive limit, and that the corresponding asymptotic characteristic is of the form

$$I\propto V_s^{\beta/\beta-1}. \tag{36}$$

The asymptotic characteristic may therefore be steeper than a quadratic, though what the precise value of β should be is difficult to determine, and may not be the same for every diode, since multicontacts may also affect it.

We have thus determined that γ approaches $1/1+b$ for high injected concentrations, but we have been able to predict only qualitatively the shape of the diode characteristic in the region of near approach. At best we may say that the characteristic should be at least quadratic in this region. Exponents as high as 3 should not be surprising, however. On the other hand, if in a given case there is no considerable "power law" portion of the characteristic, except for an eventual linear portion, we may conclude that the barrier is too low to allow injected concentrations comparable to the undisturbed majority concentration.

III. VERY HIGH CURRENTS

It must be emphasized that (37) expresses the asymptotic dependence of current upon voltage which would obtain if the equations used to describe the high current situation were valid indefinitely. There are at least two reasons why these equations become invalid for very high currents: the barrier voltage has a limiting value equal to $(\phi_b-\phi_c)$; the temperature in the region of the contact increases with current. We shall consider briefly and qualitatively how these factors should affect the diode characteristic.

As the voltage across the barrier becomes approximately equal to $(\phi_b-\phi_c)$, almost all the electrons in the conduction band of the semiconductor have sufficient energy to surmount what remains of the barrier. Consequently, the current is large enough to alter the Boltzmann distribution appreciably, and formula (24) for the electronic current no longer holds. Similarly, the concentration of holes just beyond the barrier no longer depends only on the Fermi level in the metal, as was assumed in the derivation of (6); this is because the rate at which holes flow away from the metal-semiconductor interface becomes comparable to the total rate of random entry of holes from the metal. The last fact follows from the rough equality of the hole and electron concentrations, and the similar magnitude of the two currents. Thus a "large current" (comparable to the current of random thermal agitation, or, in other words, involving drift velocities comparable to thermal velocity) of electrons implies a "large current" of holes.

We shall not here attempt an analysis for very high currents. Nevertheless we can draw one conclusion regarding the behavior of γ. The concentration of holes just beyond the barrier has an upper limit if the electronic barrier remains fixed in height. This implies that γ must drop below the value $1/1+b$ at very high currents. One would thus expect the V–I characteristic to become more nearly linear.

Eventually the region in the vicinity of the contact becomes heated significantly above the ambient temperature. The number of current carriers consequently increases, and the resistance of the diode decreases sharply. The temperature rise at the contact can be roughly estimated by assuming for simplicity that all of the heat generated in the diode is generated at the junction of metal and semiconductor and flows radially into the semiconductor by thermal conduction. Such a calculation shows that the temperature rise may reach several degrees at less than one volt forward bias for germanium of typical purity. Thus it may be necessary to resort to pulse techniques to obtain experimental isothermal diode characteristics for comparison with the theoretical results of the preceding section.

IV. INTERMEDIATE CURRENTS

It was shown in the introduction that there should be no deviation from the classical diode equation,

$$I=I_0[\exp(qV/kT)-1], \tag{37}$$

due to hole injection at low voltages. It is desirable to

know for what range of applied voltage this equation is theoretically valid.

In the one-carrier case, provided the barrier is everywhere of the same height, no deviation from the classical characteristic is expected until the voltage across the spreading resistance becomes significant as compared to that across the barrier. That one should expect the same behavior in the two-carrier case is evident from the following considerations: the electron current across a uniform barrier should have the classical exponential dependence on the barrier voltage V_b that is exhibited in Eq. (3), plus a positive correction due to the increase in carrier concentration, as represented in Eq. (24); the hole current, insofar as it is diffusion current, depends linearly on Δp, which has the same exponential dependence on the barrier voltage; the hole current due to conduction constitutes a positive correction. It is clear, then, that the dependence of the total current upon barrier voltage is as pronounced, or more pronounced, than the classical exponential dependence, and that consequently, when V_s is negligible, the form of the diode characteristic is the classical one. The forward characteristic is not steeper than the classical one, since the presence of hole conduction current requires a voltage across the spreading resistance, and the extra current will be small as long as V_s is small. Similarly, it can be deduced from the following that appreciable numbers of extra electrons at the barrier cannot exist unless V_s is appreciable.

By lowering the barrier to semiconductor electrons, the ratio of barrier resistance to spreading resistance may be made as small as desired, and hence the upper limit to the validity of (38) may be made as low as desired. The opposite proposition is also true for the one-carrier case: by raising the barrier, the ratio of resistances may be made as large as desired, and the limit of validity of the classical formula may be extended to voltages as high as desired. This proposition is not true for the two-carrier case, however, since hole injection keeps the barrier resistance from becoming infinitely large. In fact, Eqs. (8) and (9) show that in the limit of an infinitely high barrier ($I_{0n}=0$)

$$R_d/R_s=n_0/p_0, \tag{38}$$

where R_d is the total resistance of the diode for low currents, and R_s is the initial spreading resistance for $\gamma=1$, given by Eq. (19). As current is drawn, the resistance of the barrier decreases, and Eq. (37) will eventually cease to be valid.

To find a voltage constituting an upper limit on the validity of Eq. (37) we employ Eqs. (6) and (16). It will be seen that Δp in (16) has a maximum value for a given V_s when γ is unity. Rather than equating (16) and (6), we use this fact to justify the inequality:

$$n_0(e^{\alpha V_s}-1)\geq p_0(e^{\alpha V_b}-1). \tag{39}$$

The equality sign is appropriate only if γ is unity.

For barrier voltages much larger than α^{-1}, one can deduce from (39) that

$$e^{\alpha(V_s-V_b)}>p_0/n_0. \tag{40}$$

Let a voltage V_i be defined such that

$$V_i=\alpha^{-1}\log(n_0/p_0). \tag{41}$$

Then

$$V_s>V_b-V_i. \tag{42}$$

It is consequently true that

$$V_s>V_b(n-1)/(n+1),$$

when

$$V=nV_i.$$

In other words, the voltage across the spreading resistance is comparable to that across the barrier when the total applied voltage is significantly greater than V_i.

Since the voltage V_i indicates the degree to which a semiconductor is n-type (or, with an opposite sign convention, p-type), we shall hereinafter refer to it as the *indicial* voltage. It is just that voltage which, when applied to the barrier, produces an increased hole concentration Δp equal to the equilibrium electron concentration n_0. Consequently, this voltage divides the region of low currents (quasi-classical behavior) from that of high currents (quasi-quadratic behavior).

The indicial voltage for n-type germanium having a resistivity of five ohm centimeters at room temperature is approximately 0.1 volt. Thus for relatively pure samples one does not expect the classical diode characteristic to hold over a very wide range. If the barrier is low enough so that γ is considerably less than unity at low voltages, the range of validity will be further contracted. Typical values of R_{sn} and R_d for contacts made to five ohm centimeter germanium may be several hundred ohms and several thousand ohms, respectively, so that V_s may never be quite negligible.

By increasing the impurity content of the semiconductor, and thereby decreasing the number of minority carriers, it should be possible to extend the range of exponential behavior of the diode. However, early experimenters employed extremely highly n-type material, and yet found a failure of the diode law at voltages considerably less than those for which the spreading resistance should play an important role.[5] As mentioned in the introduction, this behavior has been explained on the basis of the multicontact hypothesis.

In conclusion, we point out that for extremely high barriers, the indicial voltage, though it marks an upper limit of validity of the classical diode law, need not inaugurate the power law portion of the diode characteristic, since the flow of electrons over the barrier may be effectively prohibited. In such a case the hole injection efficiency γ will remain approximately equal to unity over a prolonged voltage range. If the applied voltage is twice the indicial voltage or greater, it will be divided approximately equally between barrier and spreading resistance. Accordingly, the diode will continue to rectify exponentially, but with an exponent

equal to $\alpha V/2$ rather than to αV. The "effective alpha" for a perfect hole emitter therefore decreases to $\alpha/2$ at high voltages.

V. MEASUREMENT OF DIODE PARAMETERS

A diode of the type considered here is characterized by three parameters: the area or radius of the contact, the resistivity of the material, and the height of the barrier to electrons. It is desirable to express the principle diode quantities in terms of these parameters. Three important quantities, characterizing the semiconductor alone, are expressible in terms of the resistivity, as shown below for n-type material ($n_0 \gg p_0$). Verification of these relations is left to the reader.

$$\frac{n_0}{p_0}=\left|\frac{(b+1)\rho_i}{b\rho}\right|^2, \tag{43}$$

$$V_i=\frac{2}{\alpha}\log\frac{(b+1)\rho_i}{b\rho}, \tag{44}$$

$$\phi_e=\frac{\phi_g-V_i}{2}+\frac{3}{4\alpha}\log\frac{m_n^*}{m_p^*}. \tag{45}$$

In the above expressions ρ_i is the intrinsic resistivity, ϕ_g is the width of the energy gap considered as a potential difference, and m_n^* and m_p^* are the effective masses for electrons and holes, respectively. The current I_{0n} has already been expressed in terms of the diode parameters. The corresponding hole current constant, I_{ep}, can be written in terms of the resistivity and the area of the contact:

$$I_{ep}=\frac{(2\pi S)^{\frac{1}{2}}b\rho}{\alpha(b+1)^2\rho_i^2}. \tag{46}$$

For the plane circular contact the proper numerical factor is $4/\pi^{\frac{1}{2}}$, rather than $2\pi^{\frac{1}{2}}$. It is interesting to note that I_{ep} varies as the square root of the area, S, of the contact, while I_{0n} depends linearly upon the area. Therefore, the hole injection efficiency at low voltages should improve as the contact area is reduced.

We shall now outline, very briefly, a way of measuring the diode parameters by electrical means.

If a pulse consisting of a current step function is applied to the diode in the forward direction, the voltage across the diode will rise to a maximum, and then fall, in a time on the order of 10^{-7} seconds, to a smaller constant value. This phenomenon is due to the lessening of the spreading resistance as holes enter the semiconductor. If the current pulse is made large enough so that the associated voltage considerably exceeds the height of the barrier to semiconductor electrons, the peak voltage will be a linear function of the current, with slope equal to R_{sn}. A measurement of the resistivity then allows one to deduce the area of the contact, and to calculate I_{ep}. In conjunction, the diode resistance at low currents gives one the value of ϕ_b, provided that γ_0 is small, although this value of ϕ_b may not be meaningful in the case of multicontacts. Direct measurement of the contact area and thermal activation energy measurements of the effective barrier height at low voltages yield valuable supplementary evidence.

An apparently trustworthy measurement of the hole injection efficiency, γ, at voltages high compared to the barrier to semiconductor electrons can be obtained from the pulsing experiments, if the repetition rate is slow enough to make heating effects negligible. The actual steady-state voltage across the spreading resistance can be found from the voltage plateau occurring for later parts of the pulse; consequently one can calculate γ, using Eq. (18). Preliminary measurements by R. Rutz of the I. B. M. Research Laboratory indicate that γ does in fact approach $1/1+b$ as the current is increased.

VI. SUMMARY AND CONCLUSIONS

From the above discussion of the point contact forward characteristic, which has been based upon a model simplified by the assumptions of uniform barrier height, long lifetime, and idealized geometry, the following conclusions emerge:

(a) The expected forward diode characteristic at low voltages is the classical exponential one.

(b) There exists a voltage, characteristic of the bulk semiconductor, which constitutes an upper limit to the validity of the classical formula. This characteristic voltage is the voltage required to produce an injected hole concentration equal to the equilibrium electron concentration.

(c) The ratio γ of hole current to total current should tend to approach the fraction $1/1+b$ as the voltage is increased, and should approach it closely if the barrier is high enough to allow large injected concentrations, but not so high as to prevent effectively the flow of electrons for the voltage range of interest.

(d) As γ approaches $1/1+b$, the diode current should become approximately quadratic in the diode voltage.

It is generally agreed that γ is typically quite small at low voltages. Conclusion (c) therefore implies a γ increasing with current. A tendency of γ to increase with current has been experimentally observed by Brattain.[13]

The conclusions drawn here concerning the hole injection efficiency are not valid when there are other electrodes present. A strong sweeping field, for example, may tend to increase γ, since it diminishes the concentration of holes near the emitter, and therefore augments the hole diffusion current. Experimental evidence for this dependence of γ on sweeping field has also been obtained by Brattain.[13]

[13] Experiments by W. H. Brattain, as reported by Shockley, *Electrons and Holes in Semiconductors* (D. Van Nostrand Company, Inc., New York, 1950), first edition, p. 346.

In conclusion, the author wishes to point out that the simplified model adopted here does not include the existence of a surface p-type layer, which may be characteristic of n-type material, nor the presence of a p-type region beneath the contact, thought to be introduced by the forming process applied to commercial diodes. The latter process changes the nature of the barrier, which may become much longer than a mean free path. However, neither of these possibilities is expected to affect the high-current situation very essentially.

ACKNOWLEDGMENTS

The author wishes to express his appreciation to L. P. Hunter and to G. Knight for the discussions which suggested the problem treated here, and to E. J. Huibregtse, R. F. Rutz, E. L. Peterson, and B. E. Montgomery for their supporting experimental work. He also wishes to thank Professor L. Brillouin and G. L. Tucker for kindly reading the manuscript, and for suggesting improvements.

APPENDIX I

The derivation of (14) and (15) from (10) and (11) is quite simple. Multiply (10) by b and subtract it from (11). The result

$$-b_- i = -b\mu q[n_-\nabla\Psi - 2\alpha^{-1}\nabla p_1]$$

integrates immediately to (15).

Next, multiply (10) by $1/\gamma$ and (11) by $1/(1-\gamma)$. Both expressions are then equal to i. Subtracting (10) from (11) then leaves a relation between Ψ and p, which also can be simply integrated, yielding (14).[14]

APPENDIX II. EXACT SOLUTION FOR $\gamma_0=(1+b)$

An exact solution is valuable as a check on approximation methods, even though the case considered is somewhat special. For that reason we include a discussion of the case $\gamma_0=\gamma_e$.

[14] R. C. Prim has also obtained this solution, in somewhat different form, as part of a general discussion of carrier flow in the bulk material (R. C. Prim, III, Bell System Tech. J. **30**, 1174 (1951)).

Let us suppose that γ remains equal to γ_0. Then Eq. (16) tells us that

$$\Delta p/n_0=\alpha V_s/2.$$

Substituting in (26), we have

$$I=\frac{1+b}{2b}I_{0n}\frac{n_0}{p_0}\alpha V_s\left(1+\frac{\alpha V_s}{2}\right).$$

This is identical to (23), since for

$$\gamma_0=1/(1+b)$$

it is true that

$$\phi=I_{0n}(n_0/p_0)\alpha R_{sn}=1.$$

It follows that $\gamma=\gamma_0$ is a solution. Let

$$\beta=\frac{1+b}{2b}I_{0n}\frac{n_0}{p_0};$$

then

$$I=\alpha\beta V_s\left(1+\frac{\alpha V_s}{2}\right)$$

$$=2\beta\frac{p_0}{n_0}(e^{\alpha V_b}-1)\left(1+\frac{p_0}{n_0}e^{\alpha V_b}\right), \tag{46}$$

the second form following from

$$\frac{\alpha V_s}{2}=\frac{\Delta p}{n_0}=\frac{p_0}{n_0}(e^{\alpha V_b}-1).$$

Solving the first equation of (46) for V_s in terms of I, and setting V_b in the second equation equal to $V-V_s$, we obtain

$$I=2\beta\frac{p_0}{n_0}\{\exp[\alpha V+1-(1+2\beta^{-1}I)^{\frac{1}{2}}]-1\}$$

$$\times\left\{1+\frac{p_0}{n_0}\exp[\alpha V+1-(1+2\beta^{-1}I)^{\frac{1}{2}}]\right\}. \tag{47}$$

This formula shows the transition from an exponential to a quadratic form.

Boundary Disturbances in High-Explosive Shock Tubes*

R. G. Shreffler and R. H. Christian
Los Alamos Scientific Laboratory, Los Alamos, New Mexico
(Received June 26, 1953)

High-velocity disturbances are observed to propagate along the walls of a high-explosive operated shock tube in advance of the plane shock. Experiments are presented which determine the dependence of the geometry, energy, and velocity of the disturbance on such variables as the gas contained in the shock tube, the shock strength, and the roughness and composition of the supporting boundary. A model is constructed to explain the flow within the disturbance. Arguments are presented which show the disturbance to result from radiation originating in the luminous plane shock.

A SHOCK generated by a bursting diaphragm in a conventional shock tube will develop into a plane shock front soon after its initial formation. For the very strong shocks developed in a shock tube by the detonation of a high-explosive charge, violent disturbances may proceed down the walls well in advance of the plane shock front. The purpose of this paper is to present the results of initial investigations made in an effort to describe this phenomenon.

EXPERIMENTATION

A number of experiments which expose certain properties of the disturbance are presented in this section. The explosive system used consisted of a high-explosive lens[1] which simultaneously detonated an 8-in. square face of a 2-in. thick block of Composition *B* explosive.[2] Two types of experiments were employed to study the strong gas shock generated by this explosive system.

The setup for a type I experiment (Fig. 1) consisted of an 8-in. diameter cardboard tube, 16 in. long. This tube was closed on one end by the high-explosive system and on the other by a 1-mil thick Dural foil. The interaction of the shock with this foil was photographed by a 25 lens framing camera.[3] The necessary illumination was supplied by a long duration argon flash. An assembly including the tube is seen on the right in Fig. 1.

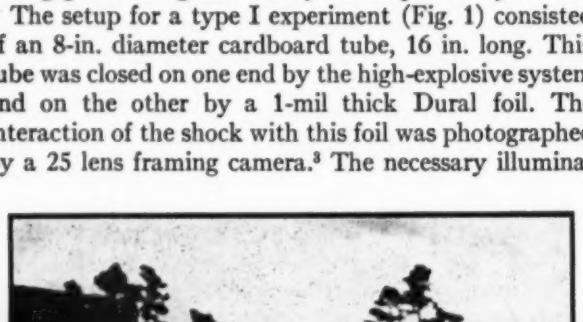


Fig. 1. A typical setup for type I experiments.

The type II experiment (Fig. 2) was designed to photograph a side view of the gas shock. A rectangular shock tube was constructed of a $\frac{1}{8}$-in. sheet Lucite with a 5 in.×5 in. cross section and 16 in. in length. The setup was aligned so that the framing camera viewed a 5 in.×16 in. side of the box normally. Since sufficient light was supplied by the radiation from the shocked gas, no additional illumination was required.

The experiments described in the following text are grouped according to purpose. For a given experiment the above descriptions are modified as noted. With the exception of Experiment I, the gas filling the shock tubes was argon.

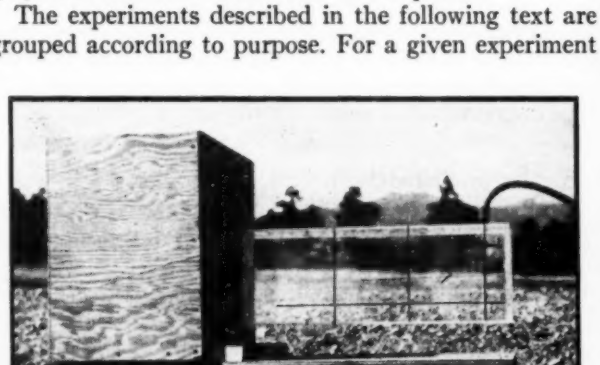


Fig. 2. A typical setup for type II experiments.

Experiment I

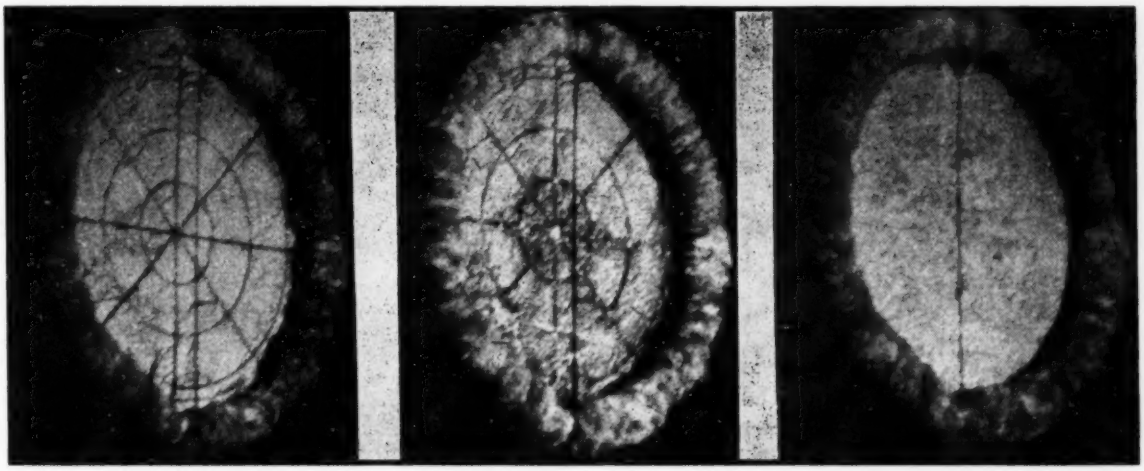
Several type I experiments were performed to study the wall disturbance in various gases. Figure 3 shows the interactions observed in this type of experiment when the shock tube was filled with argon. This experiment was repeated with various gases: air, sulfur hexafluoride, helium, butane-propane, 50/50 nitrogen-argon, and chlorine. No boundary disturbance was evident in the air and butane-propane experiments; further the shock fronts are very irregular. The helium experiment showed a plane shock with no evidence of the disturbance. The sulfur hexafluoride experiment showed an

* A version of this paper was given at the meeting of the Hydrodynamics Section of the American Physical Society held at Salt Lake City in June, 1952.

[1] J. H. Cook, Research **1**, 474 (1943).

[2] A mixture of TNT and RDX.

[3] See B. Brixner, J. Soc. Motion Picture Television Engrs. **59**, 503 (1952).



(a) (b) (c)

FIG. 3. These pictures are from a record of a type I experiment in which the shock tube had been filled with argon. Picture (a) shows the otherwise undisturbed Dural foil 8 μsec after it has been penetrated by the wall disturbance. The deterioration of the foil following the arrival of the plane shock front is recorded in the following pictures. The interval between pictures is 0.9 μsec. The dark vertical line which persists on the last picture is a reference wire 1 in. from the foil surface.

irregular shock front and a slight indication of a disturbance. The record of the 50/50 nitrogen-argon experiment showed a well-developed disturbance which preceded the plane shock by a much shorter interval than that measured in the argon experiment. Chlorine behaved in much the same fashion as argon: the shock front was plane and the disturbance was quite prominent.

It is notable that the disturbance forms only in those gases which become highly luminous under these experimental conditions. This fact was supported by the absence of a disturbance in argon-filled shock tubes when the shock velocity was lowered below 3.6 mm/μsec, the velocity below which significant radiation could not be detected.

Experiment II

Data on the trajectories of the wall disturbance and plane shock were obtained from records of type II experiments. A frame from a typical record is shown in Fig. 4. The space-time plot obtained from the entire record is shown in Fig. 5. The plane shock was found to have a constant velocity U. The wall disturbance formed at a distance x_1 along the shock tube and propagated for a distance (x_2-x_1) at a constant velocity V_1. At x_2 it decelerated to a lower constant velocity V_2. These features vary when the shock velocity is reduced: the velocities V_1 and V_2 diminish (see Fig. 11) and the distance x_1 increases. Further experiments show the dependence of this distance x_1 on the cross sectional area of the tube to be important for shock tubes of radius less than 2 in., e.g., shock tubes with radius of 2 in., 3 in., and 4 in. resulted in the same value of x_1. Reducing the radius to 1 in. approximately doubles the distance.

Experiment III

The disturbance was found to propagate along rods and wires placed parallel to the axis of the shock tube. In general the difference between this axial disturbance and the wall disturbance described in the preceding experiments was slight.

Experiments performed with a wire tilted at an angle θ to the tube axis indicated no significant change in the properties of the axial disturbance when θ was less than 20°. As θ was increased above 20° the disturbance deteriorated, so that for $\theta>26°$ no disturbance was observed.

A type II experiment was performed to observe the motion of the axial disturbance as it propagated beyond the end of a wire. 20 pictures from the resulting record

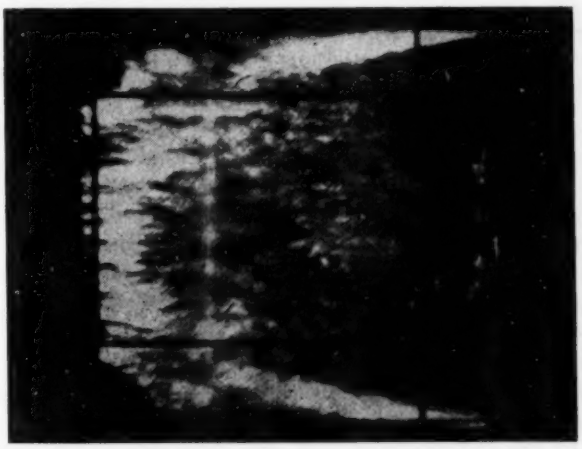


FIG. 4. This typical frame from a type II record, shows a well-formed wall disturbance in argon. The plane shock is visible at the base of the black arrow which indicates the direction of shock propagation. The black horizontal and vertical reference lines are spaced at 3 in. and 4 in., respectively.

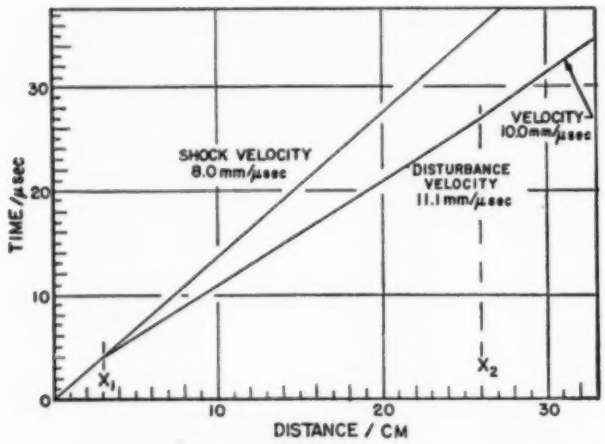


FIG. 5. The trajectory of the disturbance associated with a shock in argon. The experimental points are not shown, but lie within the thickness of the plotted lines.

are shown in Fig. 6. The velocity of the head at the disturbance after it left the wire was constant at 6.3±0.2 mm/μsec. The transverse expansion of the axial disturbance was constant at 1.8±0.15 mm/μsec.

A similar experiment was performed with a 5-in. wide, $\frac{1}{16}$-in. thick Lucite fin extending for 6 in. from the high explosive in the horizontal median plane of the Lucite box. The velocity of the disturbance, after propagating beyond the fin, was constant at 7.0±0.2 mm/μsec.

Another type II experiment was performed in which the wire extended axially 8 in. from the end of the Lucite box toward the high explosive. Thus the first 8 in. of the shock propagation were undisturbed by the axial wire. Contrary to the results obtained with the wire extending to the high explosive, no delay time was observed in the formation of the disturbance. In fact the plotted data indicate this formation to begin just before the plane shock front reaches the wire. During an interval of 7±2 μsec its geometry develops from a blunt-nosed cylinder to the characteristic cone. The initial velocity of 13.6±0.5 mm/μsec then changes to 11.6±0.4 mm/μsec, the propagation velocity of a disturbance on a wire extending to the high explosive.

Experiment IV

Two type I experiments were performed to demonstrate the energy of the disturbance. The setup was prepared with a $\frac{1}{16}$-in. diameter wire extending along the axis of the shock tube. For the first experiment a 3-in. diameter, 0.040-in. thick Dural plate was centrally attached to the one-mil Dural foil. The framing camera record showed the axial disturbance to penetrate this plate 6±1 microseconds prior to the arrival of the plane shock.

In the second experiment, the Dural plate thickness was increased to 0.127 in. Although the disturbance was unable to penetrate this target, it was able to force the plate into a dish shape, the center extending beyond the edges by 0.08 in.

The penetration properties of the disturbance as exposed by these experiments may be considered as roughly comparable to the effects produced by a 0.22 caliber long rifle lead bullet fired at close range.

The dependence of the energy of the disturbance upon the diameter of the wire became noticeable only when the wire diameter was reduced to the order of 5 mils. Disturbances propagating along such fine wires, though somewhat faster, were noticeably less energetic.

Experiment V

Two series of experiments indicated the disturbance to be composed of the gas into which it propagates. In the first series, explosive gases were excluded from the system by inserting Dural plates between the explosive system and the base of 17-mil paper shock tubes. The disturbance persisted.

The second series of experiments was performed[4] using a moving image spectrograph. Time-resolved spectrograms of the disturbances were obtained which were characteristic of the gas filling the experimental tube. Lines characteristic of the material composing the wall were barely detectable and then only when the radiation close to the surface of the wall was viewed. A spectrum characteristic of the high-explosive gases appeared later, after an interval which varied with experimental conditions.

Experiment VI

A number of experiments was performed to determine the dependence of the properties of the disturbance upon the composition and surface character of the supporting boundary.

The first experiments were type I and investigated the velocity dependence. Three rods: a 0.10-in. diameter tinned copper wire, a 0.122-in. diameter birch dowel, and a 0.126-in. diameter threaded brass rod extended the length of the tube parallel to its axis and were evenly spaced to intersect the test foil on its 4-in. diameter polar circle. The photographic record showed the average disturbance velocity on all three supporting elements to be essentially the same, i.e., independent of the composition of the boundary.

To further investigate the dependence of propagation on the surface character of the supporting boundary two type II experiments were performed. In the first experiment, 10-mil thick brass barriers extended across the 5-in. dimension of the bottom of the shock tube. The violence of interaction of these barriers with the wall disturbances increased with barrier height. It was, however, not until a barrier 20 mils high was reached that the velocity of the disturbance was significantly affected. The second experiment was similar. It was performed with the barriers replaced by successive steps in the Lucite wall of 5, 10, 20, and 40 mils. The wall disturbance was not affected by the 3 lowest steps. Interaction

[4] Under the direction of F. A. Lucy.

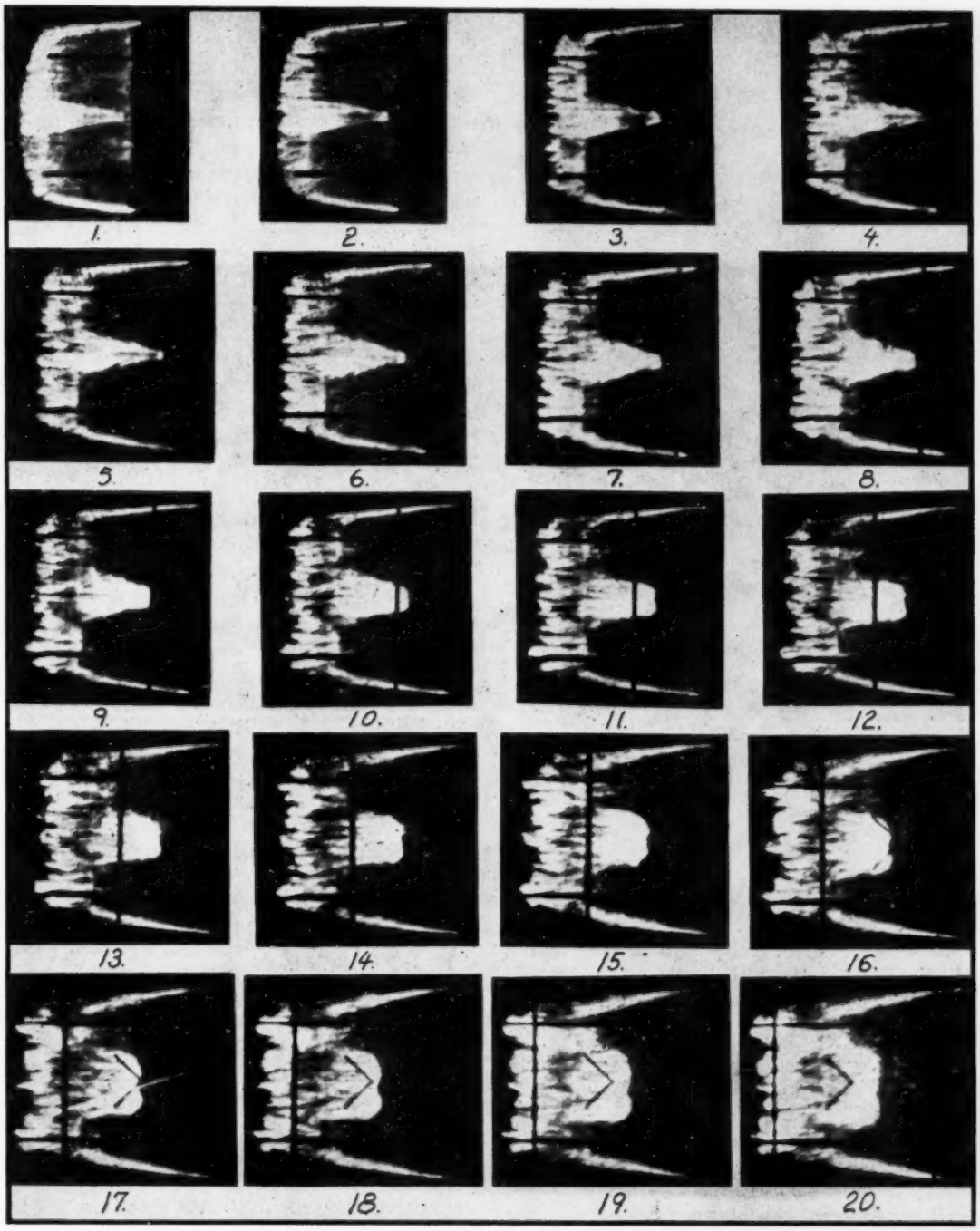


FIG. 6. These pictures partially record the recession of an unsupported axial disturbance into the shock front. The axial disturbance was formed on a $\frac{1}{16}$-in. diameter tinned copper wire which extended 10 in. from the face of the high explosive. In the first picture the disturbance has not quite reached the end of the wire. In picture 2 the effect of the terminated wire becomes evident. These pictures are spaced at 0.9 μsec intervals.

with the 40-mil step, however, was violent. The wall disturbance appeared to become turbulent.

A second type I experiment was performed to test the penetration properties of the disturbance supported by a $\frac{1}{8}$-in. birch dowel and a $\frac{1}{8}$-in. steel wire. To this end the test foil was replaced by a $\frac{1}{32}$-in. Dural plate. The photographic record indicated that the penetration properties and thus the energy associated with the disturbance to be independent of the composition of the supporting boundary.

Experiment VII

A number of previous data indicate the radiation emitted from the plane shock front to play an important

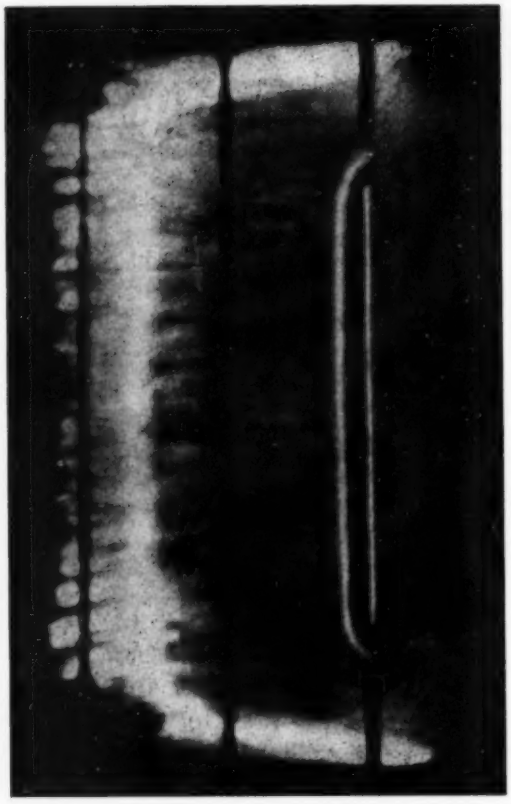


FIG. 7. The luminous secondary front. Evident in this frame are the plane shock, wall disturbance, secondary front, as well as the heated 3-in. square surface of the Lucite post.

role in the formation of the boundary disturbance.[5] A re-examination of many type II records adds to this information. In particular it was noted that, when proper film exposures were made, radiation was emitted from the sides of the shock tube well forward of the plane shock front. This radiation first appeared just forward of the plane shock front somewhat previous to the formation of the disturbance. In the frames that follow, the edge of the radiating portion of the boundary decelerates from an initial velocity in the neighborhood of 15 mm/μsec to a value approximating V_2. The wall disturbance, propagating at V_1, overtakes this edge. Within experimental error this event coincides with the deceleration of the wall disturbance to V_2.

A type II experiment was performed in which the boundaries of the forward portion of the shock tube were shielded from the radiation emitted by the plane shock front. This was accomplished by placing a $\frac{3}{4}$-mil copper foil parallel to the plane shock front in the 5-in.$\times$5-in. cross section of the experimental box 12 in. from the high-explosive surface. The foil was 4 in.$\times$5 in. and was positioned to leave a $\frac{1}{2}$-in. gap at the top and bottom of the box. The record of this experiment showed a well-formed wall disturbance approaching the foil. As the tip of the disturbance passed beyond the plane defined by the foil it decelerated to a velocity below that of the plane shock front and, in general, behaved in much the same manner as described in the experiment in which the disturbance propagates beyond the boundary of a supporting Lucite fin. This result clearly demonstrates that the existence of the preheated boundary is a necessary condition for the propagation of a boundary disturbance.

The possible association of a gaseous boundary layer, indicated in Experiment VI, with the radiation emitted by the boundary in advance of the plane shock front, led to a number of type II experiments exposing a phenomenon which we choose to call *the secondary front*. For these experiments an axial post 3 in. square and 2 in. long was attached to the far end of a 5 in.$\times$5 in.$\times$10 in. Lucite shock tube so that a 3-in. square face would locate parallel to and 8 in. from the free surface of the high explosive. The resulting photographic records show this face to become heated to incandescence, apparently by radiation emitted through the shock front. Later frames show a luminous secondary front originating at this heated surface and propagating in the argon toward the principal shock front at a velocity of 0.5$\pm$0.1 mm/μsec. This phenomenon is demonstrated in Fig. 7.

The examination of properly exposed type II records shows the specific visible radiation from the forward half of the axial and wall disturbance to be relatively weak when compared to the specific radiation from the plane shock front. This fact was confirmed by a number of smear camera experiments.

EQUATION OF STATE OF ARGON

The quantitative discussion of the boundary disturbance in argon, to be presented in this paper, requires a knowledge of the relations describing Hugoniot and adiabatic transitions in this gas.

To determine the Hugoniot, experiments were performed which simultaneously measured the shock velocity U and particle velocity u. The corresponding value of the compression μ follows from the continuity equation, and the pressure p from the momentum equation. In Fig. 8 the results of these preliminary experiments are graphically presented.[6]

For all adiabatic transitions considered in the following analysis it is assumed that equation

$$p = A\rho^{5/3}$$

is valid, where A is a function of the entropy and where ρ is the density.

THEORY

The explanation of the wall disturbance is based upon a theoretical flow model constructed in accord with the

[5] F. B. Porzel, of this laboratory, has concurrently and independently carried on investigations which are in agreement with these observations.

[6] The authors are currently pursuing this general problem of gas behavior under strong shock. Although an evaluation of the experimental techniques at the present state of development is most promising, the reader is requested to accept the data with some caution.

restrictions imposed by the experimental data. Typical geometry of the model, deduced from type II experiments, is presented in Fig. 9. From this geometry, boundary conditions are deduced which serve to define the principal flow. Attention is first given to the forward (steady state) portion of the disturbance where the angles α and β are constant, and to the initial portion of the trajectory, defined by the velocity V_1.

To an observer stationed on the tip of the disturbance, gas particles from the free stream enter the α plane with a normal velocity of $V_1 \sin\alpha$. For all values of α and V_1 obtained in these experiments this component was supersonic with respect to the undisturbed gas. It is therefore reasonable to consider the α plane to define a shock front and to make deductions concerning the state of the particles which have entered this front from the free stream. According to the equation of continuity, the expression for the normal component of the shocked particle velocity u_n is

$$u_n = \frac{1}{\mu} V_1 \sin\alpha.$$

Since the tangential component $V_1 \cos\alpha$ is preserved across the shock front, the expression for δ, the angle through which the particles are deflected, is given by the expression

$$\tan\delta = \frac{1}{\mu} \tan\alpha.$$

The experimental data presented in Fig. 9 yields a velocity for this shock front of $U = V_1 \sin\alpha = 1.74$ mm/μsec. The corresponding compression, obtained from Fig. 8, is $\mu = 3.63$. These data give a value of $\delta = 2.6° \pm 0.5°$. This value leads to the conclusion that the luminous zone, defined by the angle $(\alpha - \beta)$, is composed entirely of argon particles from the free stream; i.e., the *core* of the disturbance acts as a wedge which deflects the free stream particles along its side,

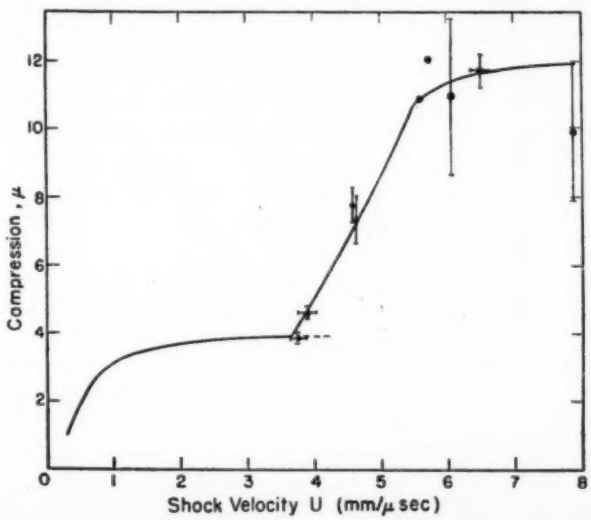


FIG. 8. Preliminary state data for argon.

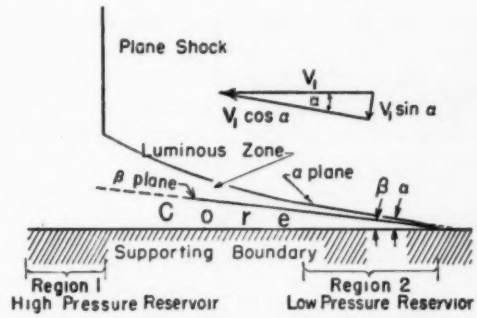


FIG. 9. This diagram illustrates the geometry of a wall disturbance associated with a plane shock velocity $U = 8.0$ mm/μsec and disturbance velocity $V_1 = 11.1$ mm/μsec. Measurements on four records all give average values of the angles α and β close to 9.5° and 7.0°, respectively, with random variation within any record as much as 2°. A lack of definition is to be expected, as the observations of these angles are made in their defining planes. Thus the film is exposed to any irregularities existing over a 5-in. depth.

and this side, the β plane, must then be a slipstream. This slipstream and the stationary position of the supporting boundary require that the particles within the forward part of the core be stationary relative to the tip of the disturbance. Thus, in the laboratory system, these particles must be moving with the velocity V_1 parallel to the supporting boundary. This requirement conforms with those experiments which display the phenomenal energy of the disturbance.

The relative pressure P within this portion of the luminous zone follows from the equation for conservation of momentum normal to the shock front

$$P = \frac{\rho_0}{p_0}\left(1 - \frac{1}{\mu}\right) V_1^2 \sin^2\alpha + 1.$$

The subscript 0 refers to local conditions, i.e., $p_0 = 0.78 \times 10^6$ dyne/cm^2 and $\rho_0 = 0.00128$ g/cm^3. Substituting the known data gives $P = 42 \pm 15$. To satisfy the conditions of steady state it is required that this pressure be maintained throughout the forward portion of the core.

The solution of the flow associated with the deflected particles in the region between the plane shock and high explosive or metal interface is difficult. It would require an assumption defining the flow configuration at the junction of the plane shock and disturbance, e.g., a 3-shock-slipstream model. This solution must then be supplemented by a subsonic compressible two dimensional solution describing the ultimate flow. For this treatment it is assumed that the deflected particles finally exist in the same state as those interacting with the plane shock. This assumption is based upon the facts that the final velocity of all argon particles in the neighborhood of the metal or high-explosive interface must be that of the interface, and that the pressure must be uniform to the extent that the disturbance has no effect upon the propagation of the plane shock.

To facilitate further discussion of the problem it is convenient to introduce certain definitions (see Fig. 9). The final state of both the deflected particles and those

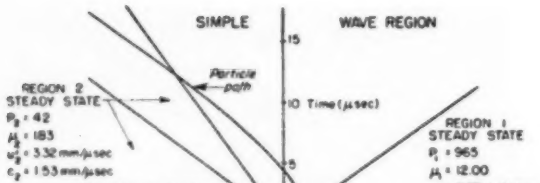


FIG. 10. The centered simple wave solution of the flow within the core.

which interact with the plane shock will be denoted by a subscript 1. This state will be referred to as region 1 or the *high pressure reservoir*. The forward (steady state) portion of the core will be denoted by a subscript 2 and be referred to as region 2 or the *low pressure reservoir*.

It remains to connect the two reservoirs by an expansion which converts potential energy in region 1 into mass motion in region 2. Since the times considered in this flow are too short for significant transport of heat, the expansion is assumed adiabatic; the high and low pressure reservoirs are conveniently connected through a centered simple wave.[7] The solution for a particular case of interest is presented in Fig. 10. The (x,t) plot of the problem is shown in the top graph. The variables defining region 1 (high-pressure reservoir) correspond to a shock velocity of 8.0 mm/μsec in argon. (Note that the selection of u_1' refers the origin to the shock front, $u_1'=u-U$.) The position of the limiting (+) characteristic which separates region 1 from the simple wave is given by the equation $x=(u_1'+c_1)t$. The expansion is terminated by the assignment of $P_2=42$. This determines the remaining variables in region 2, thus defining the disturbance head, $(V_1=U+u_2')$, and the second limiting (+) characteristic, $x=(u_2'+c_2)t$. The lower graph shows a plot of the dependent variables in space (x) at a time $t=1$ microsecond. The values of the dependent variables along the particle path, in particular u and c, determine the (+) characteristics through the equation $x=(u+c)t$. Solving this equation for $t=1$ gives the desired values of x. The plot for an arbitrary time may be obtained by multiplying the abscissa by the desired time.

[7] See, for example, Courant and Friedrichs, *Supersonic Flow and Shock Waves* (Interscience Publishers, New York, 1948), pp. 103–106.

The general dependence of the disturbance velocity V_1 on the shock velocity U may be calculated from the following equation, assuming $\gamma=5/3$,

$$V_1=u_1+0.956\left(\frac{P_1}{\mu_1}\right)^{\frac{1}{2}}\left[1-\left(\frac{P_2}{P_1}\right)^{1/5}\right].$$

This dependence for $P_2=42$, along with experimental data, is plotted in the lower curve of Fig. 11. The upper curve plots the equation for $P_2=1$, indicating the maximum velocity obtainable through an adiabatic expansion. Such a boundary condition may be approached in the formation of a disturbance on a boundary in the free stream as in experiment III.

The verification of the compression μ_2 is made indirectly from a consideration of the experiment in which the disturbance propagates in steady state along the supporting 6-in. long Lucite plane; as the tip of the disturbance reaches the end of the plane, its velocity suddenly changes from 11.1±0.15 to 7.0±0.2 mm/μsec. An approximate solution of this system is based on a one dimensional model. The system may be considered to consist initially of 2 states: state 0 defining the undisturbed gas and state 2 defining the low pressure reservoir. As state 2 propagates beyond the end of the supporting wire it is assumed to interact with state 0. This interaction would produce two shocks, one propagating back into the disturbance, the other propagating forward into the 0 state. Both shocks would be supported by a common state, 3, in which the pressure and compression would be greater than in either states 0 or 2. This problem may be mathematically expressed by the equation

$$V_1=\left(\frac{p_0}{\rho_0}\right)^{\frac{1}{2}}\left\{\left[(P_3-P_2)\frac{\mu_3-\mu_2}{\mu_3\mu_2}\right]^{\frac{1}{2}}+\left[(P_3-1)\frac{\mu_3-1}{\mu_3}\right]^{\frac{1}{2}}\right\}.$$

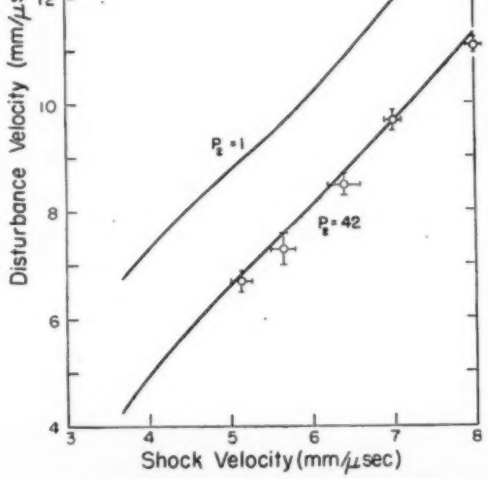


FIG. 11. The predicted disturbance velocity.

$P_3=735$ and $\mu_3=12$ are determined from the Hugoniot relations corresponding to the shock velocity (7.0 ± 0.2) at which the disturbance propagates beyond the supporting fin. The values of $P_2=42$ and $\mu_2=1.83$ are those calculated from the adiabatic expansion model (Fig. 10) for a plane shock velocity of 8.0 mm/μsec. These data give a value of $V_1=10.8\pm0.15$ mm/μsec. The discrepancy between this and the measured value ($V_1=11.1$ mm/μsec) is expected because of the transverse expansion of the disturbance. This is further demonstrated by the identical calculation for the disturbance propagating beyond the end of a wire, with a measured velocity of 6.3 mm/μsec. In such geometry the transverse expansion is permitted in 2 dimensions, with an expected further decrease in the predicted value of V_1.

The acceptability of this description of the flow is further confirmed by the agreement between the observed and the predicted disturbance geometry; e.g., from the calculated data presented in Fig. 10 the entire α plane of Fig. 9 may be reproduced well within experimental error.

Although the adiabatic flow model described in the previous paragraphs suitably describes the general flow within the disturbance, it fails to predict such features as the general independence of disturbance velocity and energy on the properties of the supporting boundary. In the large, these features may be explained by introducing a layer of gas which isolates the disturbance from the boundary. Experiment VI shows the thickness of this gaseous boundary layer to be approximately $\frac{1}{2}$ mm. Experiment VII indicates the origin to stem from the heating of the boundaries by radiation from the luminous shock. Further, the secondary front suggests that this layer propagates with a sharply defined front. Such a front, essentially parallel to the supporting wall forward of the tip of the disturbance, would permit the existence of a slipstream separating the boundary from the principal flow.

Evidence cited in Experiments I and VII indicates the origin of the disturbance to be closely associated with the radiant energy escaping through the shock front; e.g., no disturbance is observed unless the shock is accompanied by radiation. It is difficult to explain the existence of the secondary front or the boundary layer on any other premise. A crude prediction of the position at which the disturbance first forms was made on the assumption that a disturbance will not propagate until a given amount of energy has struck the boundary. This calculation did not consider the process at the boundary, the spectral distribution of the radiant energy, or the possible attenuation of the radiation in the gas. Reasonable agreement between the predicted and measured values of x_1 could be obtained.

The preceding treatment has not considered the deceleration of the disturbance to the velocity V_2. Two compatible solutions or their combination are plausible. The first is based on an extension of the calculation discussed in the last paragraph. Since $V_1>U$, the tip of the disturbance eventually traverses a boundary which has not been suitably heated; the deceleration to V_2 allows this condition to be met. A second solution is based on the fact that the high pressure head of the expansion (Fig. 10) proceeds back towards the high-explosive or metal interface at a high velocity ($u_1'+c_1$). Though a considerable layer of shocked gas may accumulate prior to the formation of the disturbance, this distance would be rapidly traversed by the head of the expansion. For these experiments the pressure at the interface would drop significantly with a corresponding slight increase in interface velocity. This interaction should finally result in a deceleration of the disturbance.

CONCLUSIONS

In this paper an effort has been made to present sufficient experimental data to describe the boundary disturbance. The theoretical treatment based on these data has resulted in a model describing the flow within this disturbance. Little is known regarding its fundamental cause though it is believed to be closely associated with the radiation escaping through the plane shock front, particularly as this radiation contributes to the formation of the boundary layer or secondary front. This aspect of the problem is receiving further consideration.

ACKNOWLEDGMENTS

The experiments described in the text were performed under the direction of Charles L. Proctor. The authors further acknowledge the assistance of W. E. Deal, R. E. Duff, F. A. Lucy, D. P. MacDougall, C. W. Mautz, E. W. Salmi, J. M. Walsh, and F. J. Willig.

Arcing at Electrical Contacts on Closure. Part IV. Activation of Contacts by Organic Vapor

L. H. GERMER
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received June 29, 1953)

The activation of electrical contacts by decomposition products from organic vapors depends upon an adequate surface layer of adsorbed molecules. In earlier work minimum vapor pressures were given for activation by a number of organic compounds. From new experiments reported here the time available for adsorption between successive arcs is equally important, so that the minimum vapor pressure for activation to occur is inversely proportional to the time available for adsorption, or directly proportional to the rate of operation. The effective contact area and the energy are important also and, although quantitative data have not been obtained, it is reasonable to guess that these variables are connected by a relation not far from $p/n=KE/A$, where n is the minimum rate for activation to occur for the pressure p on the effective contact area A with the available energy E. Quantitative experiments have been made only upon the activation of palladium contacts by vapor of the organic compound known as fluorene, and only the proportionality between p and n has been established.

THE occurrence of an arc between electrodes approaching each other at a low difference of potential is dependent upon the nature of the electrode surfaces and upon the electrical circuit. An arc occurs between clean noble metal surfaces only if the circuit inductance and resistance are quite low, but surfaces activated by decomposition products from organic vapors arc through much larger inductances and resistances. The two types of arcs, "inactive" and "active," are characterized by widely different minimum arc currents and by widely different gross electric fields at which the arcs strike.[1] It has been concluded that the activation process, which takes place when a circuit is closed repeatedly at a pair of contacts, consists of the adsorption of organic vapor upon the surface of one of the contacts and its decomposition there to form carbonaceous material.[2] This activation by organic vapors occurs frequently on relay contacts of the telephone plant, and in some cases results in enormously increased rate of erosion of the contact metal and a corresponding decrease of relay life. Experiments leading to insight into the activation process are therefore of practical value as well as of scientific interest.

Activation takes place only when the vapor pressure is sufficiently high, and an approximate value of the minimum pressure which will just cause activation has been given for a number of different compounds (see Table I of Reference 2). Since these tests were made it has been found that this minimum pressure varies greatly with rate of operation of the contacts, with their areas, and with the available energy. The new factors appear because an arc has other effects upon adsorbed organic matter besides its carbonization; an arc will burn off some of the material already carbonized. Whether or not there is activation must depend upon the relative rates of burning off carbon and of forming more of it. The balance between the two processes is dependent upon rate of adsorption and upon energy. Increased adsorption favors activation, but increased energy favors burning off carbon and thereby maintaining the surfaces clean or inactive. For low vapor pressures adsorption per operation can be expected to be proportional to the time the contacts are separated between closures and to the effective area, as well as to the vapor pressure. The material burned off completely should increase, perhaps not linearly, with the energy dissipated on each closure. One can imagine an arc of diameter d cleaning an area $\pi d^2/4$ but producing new carbonaceous stuff around the periphery of the arc on an area $\pi d(\Delta d)$. Contacts can be expected to become active when the ratio of these areas, $d/4(\Delta d)$, is less than unity, and to remain inactive when the ratio is greater than unity. Perhaps Δd will be independent of energy but d will surely increase with increasing energy, so that there will exist a limiting maximum energy for activation for each particular vapor pressure and rate.

This paper is concerned with a quantitative study of the relation between the minimum vapor pressure at which activation will occur and the time the contacts are separated between successive closures. Observations of a qualitative nature are reported also of the effect of area and energy upon the minimum pressure for activation.

The investigation of the relation between minimum vapor pressure and time available for adsorption between successive closures is very simple in principle. One operates a pair of contacts for a long time in some standard electrical circuit in a closed container in a constant pressure of a particular activating vapor. From a series of experiments of this type, at different rates of operation, one finds whether or not for this pressure there is a limiting rate below which the contacts will become active and above which they will not become active. If there is such a limiting rate, and if the series of experiments is repeated for a number of

[1] L. H. Germer, J. Appl. Phys. 22, 1133–1139 (1951), Table I; L. H. Germer and J. L. Smith, J. Appl. Phys. 23, 553–562 (1952), Table I.
[2] L. H. Germer, J. Appl. Phys. 22, 958–959 (1951).

different pressures of the same vapor, one obtains finally a relation between limiting vapor pressure and rate.

EXPERIMENTAL METHOD

The simplest way of maintaining a constant pressure is to place more than enough of the organic material to saturate the air in a tight enclosure containing the experimental contacts and to hold the temperature of the enclosure fixed. The organic compound must be carefully chosen in order that the temperatures which produce appropriate pressures shall be in a convenient range. It is clear, for example, that benzene would be an inconvenient activating agent for these experiments, because the vapor pressure of benzene at room temperature is about 100 mm Hg whereas the activating pressures at which experiments must be carried out in benzene are of the order of 0.1 mm Hg.

In preliminary experiments it was found that for naphthalene also the required range of pressures would require uncomfortably low temperatures. For anthracene, on the other hand, one would have to operate at quite high temperatures. These first tests led to a search for a benzene derivative having vapor pressures intermediate between those of naphthalene and anthracene. Such a compound, fluorene, diphenylenemethane, $C_6H_4CH_2C_6H_4$, was supplied by Mr. A. N. Holden. This compound has melting and boiling points intermediate between those of naphthalene and anthracene,* and it is to be presumed that its vapor pressures at various temperatures lie also between those of these two closely related compounds. It has turned out that the vapor pressures of fluorene which are required are actually maintained by temperatures in the neighborhood of ordinary room temperature, in a convenient range for experimentation. All of the tests reported here were carried out upon the activation of palladium contacts in fluorene vapor.

The enclosure used for the tests is a small glass bottle within which a pair of contacts can be opened and closed by means of an electromagnet outside the bottle. In each test the bottle is held at a temperature somewhat below room temperature by immersion in a water bath, or above room temperature by a furnace surrounding both the bottle and the magnet. Before each test the bottle and the contact assembly are separately heated in a hydrogen flame to remove organic materials, and the cleanliness resulting from this heating is tested by operating the contacts within the empty bottle at the temperature at which the bottle is to be held after placing fluorene in it. If the contacts do not remain inactive in this test, it is obvious that the cleaning has been inadequate and it must be repeated more carefully.

*

	Naphthalene	Fluorene	Anthracene
M. P.	80°C	116°C	217°C
B. P.	218°C	295°C	354°C

A NEW TEST FOR ACTIVATION

Tests for activation are very easy and straightforward when an activating vapor is present in great excess. If, however, the pressure of vapor present in a test chamber is not much more than is sufficient to cause activation, whether or not activation will actually result is very dependent upon the method of testing. In the work reported here a novel method has been used for developing activation and testing for it. A pair of palladium contacts within the test bottle is made to open and close repeatedly at a fixed rate, remaining open half of the time and closed half of the time. At each closure a small condenser is discharged through a noninductive resistance which is so low that the initial current is well above the minimum arc current for inactive electrodes. No current is broken when the contacts separate. In this circuit there is always an arc on closure and the heat of this arc can be depended upon to decompose adsorbed organic vapors, and thus to activate the contacts if enough vapor is present.

Since there is always an arc on closure in this kind of a circuit whether the contacts are active or inactive, a special kind of observation must be made to discover their condition. The method employed is based upon the fact that active contacts arc at a considerably lower field strength than do inactive contacts, and therefore at a considerably greater separation for the same potential difference. Each arc throws up a mound of metal which will weld the electrodes together if they are sufficiently close but will not weld them if they are far apart. By a suitable choice of circuit constants it can be so arranged that inactive contacts are welded on each closure but active contacts are not welded. Determining for each closure whether the contacts are welded or are not welded is very easy.

The method can succeed only when circuit constants are appropriately chosen. Data are available for making suitable choices, although some of the data are for silver electrodes, some for platinum, and some for palladium. The initial current in a closure arc is $I_0=(V_0-v)/R$ amperes, where V_0 is the initial voltage on the condenser C, v is the arc voltage, and R is the circuit resistance. The energy in the arc is $E=Cv(V_0-v-I_mR)\times10^7$ ergs, where I_m is the minimum arc current. The electrode separation at which the arc strikes is $s=V_0/f$ cm, where f is the critical field at which it strikes. The height of the mound thrown up by the arc is $s^*=BE^{\frac{1}{3}}$ cm (reference 1a, p. 1136, footnote 3). We must choose the parameters V_0, R, and C to satisfy three conditions. These are:

1. $I_0 \gg I_m$ in order that an arc will always strike and thereby supply heat for activation if sufficient vapor is present.

2–3. $s_i \ll s^* \ll s_a$, where s_i is the separation at which an inactive arc strikes and s_a the considerably greater separation for an active arc.

Using numerical values† of f and I_m for active and inactive metal, and of B and v, calculations have been made for a number of different assumed sets of circuit constants. The parameters finally chosen as suitable are

$$V_0=150 \text{ volts},$$
$$C=3\times10^{-9} \text{ farad},$$
$$R=47 \text{ ohms},$$

which give $I_0=2.9$ amp, $s_i=6500$A, $s_a=19\,000$A, and $s^*=7600$A. Although these values satisfy the three above requirements quite nicely, it should be pointed out that the numerical data upon which the calculations were made were obtained from three different metals but are nevertheless being applied to the single metal palladium. Furthermore, we have no knowledge that the equation for mound height applies both to active and inactive surfaces. The calculations must therefore be considered as leading only to rough guesses of suitable values of the circuit constants. The payoff is the actual test. It turns out that the tests are satis-

† Reference 1, p. 1136, Table I and footnote 3. Reference 1, p. 559, Table I. Reference 2, p. 957, Table II.

FIG. 1. Typical oscilloscope trace of the closure of fully active contacts. Open circuit following an arc and preceding touching of the contacts is represented here by the central spot. From the relative location of this spot the minimum arc current is calculated to be 0.13 amp.

factory, and all experiments reported in this memorandum were made with the circuit constants just listed.

In this circuit there is an arc on every closure of the contacts. If the contacts are inactive the arc is ended by closure, but if they are active there is an open immediately following the arc. One determines whether or not the contacts are active by observing on an oscilloscope whether or not there is an open preceding closure. A typical oscilloscope trace for active contacts is reproduced here as Fig. 1. On this trace the ordinates represent potential across the contacts and the abscissae potential across the condenser. One sees three spots; the lowest represents closure, zero potential across the contacts and across the condenser; the top spot corresponds to separated contacts, 150 volts across the contacts and 150 volts across the condenser also; the intermediate spot is the open circuit. For inactive contacts the initial arc ends in closure and the open circuit spot is not present. The test for activity is very simple; one simply notes whether there are *two spots* or *three spots*.

The potential corresponding to the open circuit spot of Fig. 1 is $V_1=20$ volts. The minimum arc current is $I_m=(V_1-v)/R=0.13$ amp, which corresponds to very active palladium contacts.‡ This method of testing not only tells at once whether or not contacts are active, by the presence or absence of the third spot, but if active it gives at a glance the minimum arc current. The chief advantage of the method is, however, the fact that there is plenty of heat supplied to inactive contacts to produce activation if the organic vapors are adequate.

One can see in Fig. 1 a faint trace of a horizontal line slightly below and to the right of the central spot; this is the end of the arc preceding the open circuit represented by the spot. Most of the horizontal line corresponding to the arc is too faint to be seen because the time of the arc is so very short, comparable with $CR=0.14\times10^{-6}$ sec. The intensity of the open circuit spot is proportional to the duration of the open and therefore to the difference between the separation of the contacts at which the active arc strikes and the height of the mound thrown up by the arc. The intensity of the spot thus increases with increase in activity of the contacts. When initially inactive contacts begin to become active one observes first a very high open circuit spot (high minimum arc current) of low intensity (high initial field strength and therefore small separation and very short open circuit time). As activity increases the spot moves lower along the diagonal line connecting the end spots and at the same time increases greatly in intensity. There is, of course, a great deal of statistical variation in these phenomena.

‡ The very low values down to 0.03 amp (see Table I, p. 1136, Reference 1) often found for silver contacts apparently never occur for palladium.

EXPERIMENTAL RESULTS

Initially inactive palladium contacts are operated for 50 000 closures at a fixed rate in the glass enclosure containing fluorene at a fixed temperature. The significant observation is whether or not the contacts are active at the end of this time. The results obtained from repeating this test at a number of different rates for a number of different fluorene temperatures are plotted in Fig. 2. This is a semilog plot with abscissas representing temperature on a linear scale, and ordinates rates on a logarithmic scale. The circles correspond to tests at the end of which the contacts were inactive and the crosses to tests which resulted in active contacts.§

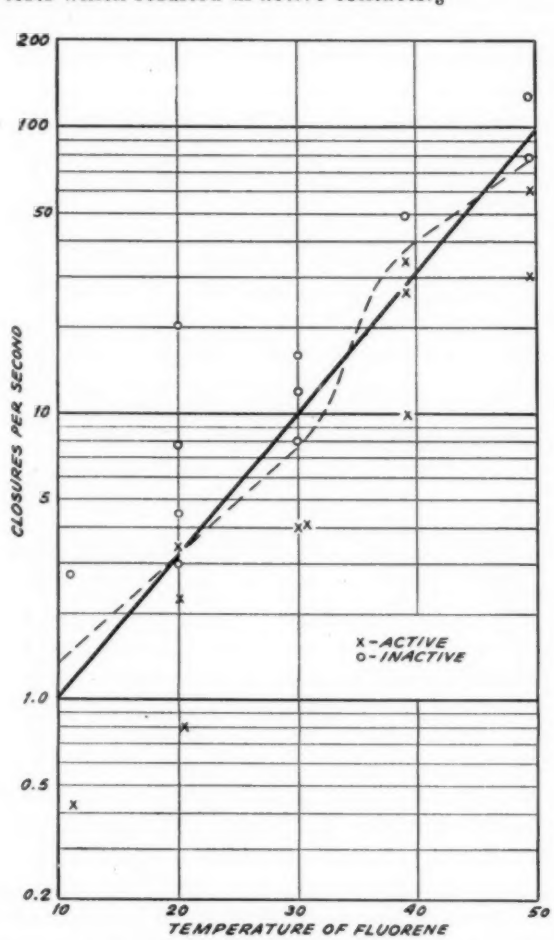


FIG. 2. Activation of palladium contacts by operating at various rates in the saturated vapor of fluorene. The crosses represent tests in which the contacts became active in 50 000 closures and the circles tests in which they were still inactive. The maximum rate of operation n for activation is represented by the solid line $\log_{10}n=-0.5+0.05(\Delta T)$ where ΔT is the fluorene temperature in centigrade degrees.

§ Many tests were also made in which the contacts were operated for 500 000 times before checking for activation. Maximum rates for activation were found to be uniformly considerably higher than for the smaller number of operations. The plot of Fig. 2 does not therefore correspond to conditions of equilibrium.

One sees from the points of Fig. 2 that there is a relation between the fluorene temperature and the maximum rate of operation n at which contacts will become active. This relation is perhaps best given by the dashed line on the figure, but the solid straight line certainly fits the data well enough. This line only will be considered in the following paragraphs.

Although vapor pressures of fluorene have never been measured one can probably guess well enough how rapidly its pressure varies with temperature. The usual expression for vapor pressure can be written

$$\log_{10}p=b-0.05223a/T_0+0.05223a(\Delta T)/T_0^2$$

for a range of values of ΔT around the absolute temperature T_0. For our present use $T_0=273$ and ΔT is centigrade degrees. The equation of the straight line of Fig. 2 is

$$\log_{10}n=-0.5+0.05(\Delta T).$$

If one assumes that n is proportional to the vapor pressure, $n/p=k$, these equations give relations between $\log p$ and ΔT which have the same slope for $a=71\,000$. In other words, if the constant a for fluorene in the general vapor pressure equation is 71 000 for temperatures near 0°C, the straight line of Fig. 2 is equivalent to the statement that the maximum rate for activation is proportional to the fluorene pressure. In Table I are given values of the constant a for the closely related compounds naphthalene and anthracene for the lowest temperature ranges for which data[3] are available:

TABLE I.

	a	Temperature range
Naphthalene	71 401	0 to 80°C
Anthracene	70 390	100 to 160°C

One sees that if the constant a for fluorene for temperatures just above room temperature is close to the value which it has for anthracene and naphthalene, the straight line of Fig. 2 does really mean quantitatively that the minimum rate for activation is proportional to the vapor pressure. The proof of this relation seems to be as satisfactory as the data permit.

Although this experiment indicates quite clearly that the maximum rate for activation is proportional to the vapor pressure, there is obviously considerable scattering in the data. From other tests it appears that the maximum rate for activation at a fixed vapor pressure depends, at least in qualitative fashion, inversely upon the available energy and directly upon the area of the contacts over which arcs can occur. It seems likely that variation of contact area, resulting, for example, from roughness due to erosion, may account for scattering of the data in the present experiments. In fact, with the contact area as little under control as it is in these tests, it is really remarkable that results were ever obtained as consistent as those indicated in Fig. 2.

There are no quantitative data relating maximum rate for activation and contact area. The general dependence of maximum rate upon area is, however, readily observable in qualitative fashion. It is found that the flat palladium surfaces of the contacts of a telephone relay are activated at vapor pressures several times less than the lowest pressure which will activate crossed wires of palladium operating at the same rate and in an identical circuit. The relation between minimum vapor pressure (or maximum rate for a constant vapor pressure) and area is not easily found because there seems to be no way of knowing the area in any particular experiment. This area is not the gross contact area but only that portion of the area on which an arc can strike. The simple ideas expressed at the beginning of this paper give a reason for believing that the maximum rate may be about proportional to the true arcing area. If this is true, the value of the quotient p/n must be inversely proportional to the area A, or $p/n=\kappa/A$.

It is observed that ease of activation depends inversely upon energy in the arcs. Contacts are activated much more easily at low energies than at high. Here again quantitative data have not been obtained, but qualitative observation of this effect is strikingly easy. The maximum rate for activation may perhaps be inversely proportional to the energy, although there is no known reason why this should be strictly so. If the relation is as simple as this we can write $p/n=KE/A$.

[3] *International Critical Tables* (McGraw-Hill Book Company, Inc., New York, 1930), Vol. 2, p. 208.

Use of Restricted Variational Principles for the Solution of Differential Equations*

PHILIP ROSEN
Applied Physics Laboratory, The Johns Hopkins University, Silver Spring, Maryland
(Received July 6, 1953)

A technique is given by which nonlinear differential equations may be solved approximately, using variational principles for which certain restrictions are made. Ordinary first- and second-order equations are discussed in detail. The method is extended to encompass partial differential equations and applied to the Boltzmann transport equation.

THE study of irreversible thermodynamics has led to the formulation of irreversible processes by means of variational principles which differ from the usual variational principles of ordinary mechanics.[1,2] In the Onsager formulation, for example, the total rate of entropy production is a minimum with respect to variation of the generalized currents when the generalized forces are held fixed or specified. In addition there are restrictions on the currents due to continuity conditions.

The fact that the generalized forces are held fixed during a variation of an integral is extremely important for it means that the variation is a restricted or partial variation. This technique has led the author to the present study because one can develop variational principles for any differential equation provided that one holds the proper quantities specified or fixed when a variation of an integral is carried out. In particular one is generally interested in solving nonlinear differential equations approximately without linearizing the equation.

A. ORDINARY FIRST-ORDER EQUATIONS

Let us consider the integral

$$I=\int_{x_1}^{x_2}\left[g(x,y)\frac{dy}{dx}+\frac{y^2}{2}f\left(x,y,\frac{dy}{dx}\right)\right]dx, \tag{1}$$

where y is given at x_2 and $g(x,y)=0$ at x_1. If we now let $\delta I=0$, subject to the restriction that $g(x,y)$ and $f(x,y,dy/dx)$ be held fixed, we obtain the Euler-LaGrange equation

$$\frac{d[g(x,y)]}{dx}=yf\left(x,y,\frac{dy}{dx}\right). \tag{2}$$

To be more rigorous, when $g(x,y)$ and $f(x,y,dy/dx)$ are held fixed we mean that $g(x,y')$ and $f(x,y',dy'/dx)$ are fixed when y is varied and that $y=y'$ is an auxiliary relationship to be satisfied after the variation is carried out. Equation (2) can easily be made to fit the form of any first-order equation. The only mathematical requirements are that dy/dx and $g(x,y)$ be continuous and that the integral exists in the inteval (x_1,x_2). If Eq. (2) is satisfied, then

$$I=I_0=\int_{x_1}^{x_2}\left[g(x,y)\frac{dy}{dx}+\frac{y}{2}\frac{d}{dx}g(x,y)\right]dx. \tag{3}$$

Consequently, if we have an approximate solution $y=F(x)$ to Eq. (2), then we can use the difference $I-I_0$ as a measure of the accuracy of the solution.

In an actual case we might express a solution to Eq. (2) as $y=G(x,\alpha)$ where α is a parameter to be determined. We then can express

$$g(x,y)=g[x,G(x,\beta)]$$

and

$$f\left(x,y,\frac{dy}{dx}\right)=f\left[x,G(x,\beta),\frac{dG}{dx}(x,\beta)\right],$$

where β is a temporary alias for α. We then evaluate the integral $I=I(\alpha,\beta)$. Now the integral I must be varied with respect to α with β fixed. Thus we calculate $\partial I(\alpha,\beta)/\partial\alpha=0$ and then set $\beta=\alpha$ after the differentiation. This approach leads to an algebraic equation for α. In general one may use as many parameters as one desires. Also one can expand the dependent variable y in a series of orthogonal functions and use the coefficients as the parameters to be varied.

Let us consider the simple equation $dy/dx=y^2$ as an example, with $y=1$ when $x=1$. The exact solution of this problem is $y=1/(2-x)$. To check our method, let us put $y=1/(2-x)^n$ and use our technique to find n. In order to make Eq. (2) the same as the equation under discussion we have a great deal of choice for $g(x,y)$ and $f(x,y,dy/dx)$. If we choose $g(x,y)=xy$, then $f(x,y,dy/dx)=1+xy$. Using m as an alias for n we have

$$g=\frac{x}{(2-x)^m},\quad f=1+\frac{x}{(2-x)^m}.$$

When we compute I using the limits $x=0$ and $x=1$, we find

$$I=\frac{n}{m+n}-\frac{n}{(m+n)(m+n-1)}[1-2^{1-m-n}]+\frac{1}{2}\left[\frac{1}{2n+m-1}+\frac{(1-2^{1-2n})}{2n-1}-\frac{(1-2^{2-m-2n})}{(2n+m-1)(2n+m-2)}\right].$$

* Supported by the Bureau of Ordnance, U. S. Navy.
[1] L. Onsager, Phys. Rev. 37, 405 (1931).
[2] S. R. DeGroot, *Thermodynamics of Irreversible Processes* (Interscience Publishers, Inc., New York), p. 196.

After calculating $\partial I/\partial n=0$, and setting $m=n$ we obtain

$$\left[\frac{2^{-2n}}{2n-1}-\frac{2^{2-3n}}{(3n-1)(3n-2)}\right]\log_e 2-\frac{(1-2^{1-2n})}{4n(2n-1)^2}(4n-1)+\frac{(1-2^{2-3n})}{(3n-1)^2(3n-2)^2}(6n-3)+\frac{9n^2-10n+1}{4n(3n-1)^2}=0.$$

The solution of this equation is $n=1$, as was expected. There are no other roots in this case. If a situation arises in which there are multiple roots and it is difficult to distinguish which root is the correct one, we can calculate the quantity $(I-I_0)/I_0$ to determine which root makes the latter quantity smallest.

B. ORDINARY SECOND-ORDER EQUATIONS

The method given in Sec. A can easily be extended to second-order equations. Consider the integral

$$J=\int_{x_1}^{x_2}\left[g\left(x,y,\frac{dy}{dx}\right)\frac{\left(\frac{dy}{dx}\right)^2}{2}+h\left(x,y,\frac{dy}{dx}\right)\frac{dy}{dx}+f\left(x,y,\frac{dy}{dx},\frac{d^2y}{dx\,dx^2}\right)y\right]dx, \quad (4)$$

where y is given at x_1 and x_2. If we let $\delta J=0$ with g, h, and f fixed we obtain the Euler-LaGrange equation

$$\frac{d}{dx}\left(g\frac{dy}{dx}\right)+\frac{dh}{dx}=f. \quad (5)$$

If the boundary conditions are different, one can modify the integrand of J accordingly. For example if y is given at x_1 and dy/dx is given at x_2, we could replace the term $(dy/dx)^2$ by

$$\left[\frac{dy}{dx}-\left(\frac{dy}{dx}\right)_{x_2}\right]^2$$

and take $h=0$. If y and dy/dx are both given at x_2, then we can replace $(dy/dx)^2$ by

$$\left[\frac{dy}{dx}-\left(\frac{dy}{dx}\right)_{x_2}\right]^2$$

and pick g and h so that they both vanish at x_1. Such procedures insure the vanishing of surface integrals obtained by the usual partial integrations when the integral J is varied.

In some cases a differential equation which is nonlinear may be obtained from a variational principle for which restrictions like those above are not necessary. For example,

$$\delta\int_{x_1}^{x_2}\left[\frac{1}{2}\left(\frac{dy}{dx}\right)^2+\frac{y^3}{3}f(x)+\frac{y^2}{2}g(x)+yh(x)\right]dx=0$$

with y fixed at x_1 and x_2, yields

$$d^2y/dx^2=y^2f(x)+yg(x)+h(x).$$

However, such cases are exceptions. In general, restrictions on the way a variation is to be carried out are necessary if we are dealing with a nonlinear problem.

C. PARTIAL DIFFERENTIAL EQUATIONS

The techniques given in Secs. A and B can be extended to partial differential equations. Each type of equation will require a different approach but the basic technique will remain essentially the same. Elsewhere[3] the author has developed a variational principle for the Navier-Stokes equation. In this case the rate of viscous dissipation is a minimum when the fluid velocity is varied, the acceleration is fixed, and the conservation of mass and energy is satisfied. Another example is the problem of diffusion when the diffusivity D is a function of the concentration c.

The variation

$$\delta\int\left[\frac{D(c)}{2}(\nabla c)^2+c\frac{\partial c}{\partial t}\right]dx\,dy\,dz=0,$$

with $D(c)$ and $\partial c/\partial t$ fixed and c specified on the boundary, yields the equation

$$\nabla\cdot[D(c)\nabla c]=\partial c/\partial t.$$

The above methods prove to be useful in solving the Boltzmann transport equation of kinetic theory. In some situations the usual expansion method of Chapman and Enskog fails. In particular, when gas flows through a shock wave the velocity distribution function changes appreciably in a distance comparable with the mean free path.

Let us consider the integral

$$I_B=\int_\tau\int_v\left\{f'\mathbf{v}\cdot\nabla f+f'\mathbf{a}\cdot\nabla_v f+\left[\left(\frac{\partial f'}{\partial t}\right)_{\text{collisions}}-\frac{\partial f'}{\partial t}\right]f\right\}d\tau\,d\mathbf{v}$$

where $f(x,y,z,v_x,v_y,v_z)$ is the velocity distribution function, ∇_v is the gradient in velocity space, $\mathbf{a}$ is the acceleration of a molecule due to external forces, and f' is a temporary alias for f(i.e., $f'=f$ after the variation is carried out). If we vary f holding f' fixed we then have

$$\delta I_B=\int_\tau\int_v-\left[\nabla\cdot f'\mathbf{v}+\nabla_v\cdot f'\mathbf{a}-\left(\frac{\partial f'}{\partial t}\right)_{\text{collisions}}+\frac{\partial f'}{\partial t}\right]\delta f\,d\tau\,d\mathbf{v}+\int_v\int_S f'\mathbf{v}\cdot d\mathbf{S}\,\delta f\,d\mathbf{v}+\int_\tau\int_{S_v}f'\mathbf{a}\cdot d\mathbf{S}\,\delta f\,d\tau,$$

[3] P. Rosen, J. Chem. Phys. **21**, 1220 (1953).

where S is the space surface enclosing the volume τ and S_v is the velocity surface at infinite velocity. If we take f to be specified on S, then since f must vanish at infinite velocities, both surface integrals vanish. If we now require $\delta I_B=0$, then

$$\int\int\left[\frac{\partial f'}{\partial t}+\mathbf{v}\cdot\nabla f'+\mathbf{a}\cdot\nabla_v f'-\left(\frac{\partial f'}{\partial t}\right)_{\text{collisions}}\right]\delta f d\tau d\mathbf{v}=0.$$

If we now let f' approach f, we obtain the Boltzmann transport equation. In an actual problem the limiting process of letting $f'\rightarrow f$ after the variation is easily accomplished by using a parametric representation.

In choosing a trial distribution function one must recall that the conservation of energy, matter, and momentum must be satisfied and thus the chosen distribution function must satisfy the equation

$$\int\Phi\left[\frac{\partial f}{\partial t}+\mathbf{v}\cdot\nabla f+\mathbf{a}\cdot\nabla_v f\right]d\mathbf{v}=0,$$

where Φ represents any of the collision invariants.[4] This restriction will yield relations between the parameters in terms of which the distribution function is expressed.

When the integral,

$$\int\int_\tau\int_v\left\{f\mathbf{v}\cdot\nabla f+f\mathbf{a}\cdot\nabla_v f+\left[\left(\frac{\partial f}{\partial t}\right)_{\text{collisions}}-\frac{\partial f}{\partial t}\right]f\right\}d\tau d\mathbf{v}dt,$$

equals the integral

$$2\int\int_\tau\int_v\{f\mathbf{v}\cdot\nabla f+f\mathbf{a}\cdot\nabla_v f\}d\tau d\mathbf{v}dt,$$

we have found the correct solution. If the integrals are not equal, their difference can be used as a measure of the accuracy of the assumed distribution function.

[4] Chapman and Cowling, *The Mathematical Theory of Non-Uniform Gases* (Cambridge University Press, Cambridge, 1939).

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 3 MARCH, 1954

Apparent Density of Thin Evaporated Films*

M. S. Blois, Jr.,† and L. M. Rieser, Jr.‡
Department of Physics, Stanford University, Stanford, California
(Received July 8, 1953)

Measurements of the apparent densities of copper and silver films formed by evaporation are described. The interpretation of the results and the manifestations of low apparant densities in physical experiments are discussed.

INTRODUCTION

THE direct observation of thin evaporated films made possible by the electron microscope has by now completely confirmed the conjectures of earlier workers who by studying the effect of film thickness on the resistivity of metal films concluded that an agglomerated or particulate state preceded the continuous one as film thickness was increased. Picard and Duffendack,[1] Levinstein,[2] and Hass and Scott,[3] have in their papers indicated some of the parameters, both of materials and technique, on which the structure of evaporated films depends.

Since many of the unique features of thin film behavior, including the apparent density, arise from processes at the substrate-film interface, the nature of the substrate surface becomes of considerable importance. If the surface has been carefully cleaned before evaporation so that the oncoming evaporated atoms are strongly bound into the potential wells of the surface, several orienting effects may be expected. If the substrate is a crystal, epitaxy[4] may occur and the evaporated atoms will take up a periodicity associated with the underlying lattice. This will continue over such distances as can be accommodated by the interatomic spacing of the evaporated material compared to that of the substrate, i.e. until the "phase" relationship becomes poor. At this point one microcrystal will end and at the next or nearby lattice point another may begin. Similarly, with noncrystalline substrates which are highly oriented, such as a stressed film of linear organic molecules, the atoms first to condense on this surface will become preferentially bound and tend to duplicate the structure of the substrate.[5]

If the substrate is contaminated by water vapor, adsorbed gases, or oil or grease, all of which may be

* This work was assisted in part by a grant by the Sloan-Kettering Foundation.
† Now at National Bureau of Standards, Corona, California.
‡ Now at Dartmouth College, Hanover, New Hampshire.
[1] R. G. Picard and D. S. Duffendack, J. Appl. Phys. 14, 291 (1943).
[2] H. Levinstein, J. Appl. Phys. 20, 306 (1949).
[3] R. S. Sennett and G. D. Scott, J. Opt. Soc. Am. 40, 203 (1950).
[4] G. W. Johnson, J. Appl. Phys. 21, 449 (1950).
[5] M. S. Blois, Jr., Science 114, 175 (1951).

considered amorphous on the scale of observation provided by electron microscopy, then there appear no observed ordering effects and further, the strength of the binding appears to be reduced. In the latter case the agglomeration is quite pronounced. It is possible that the binding to a contaminated surface may be greater than otherwise, and this indeed has been reported. This could only happen, however, if the substrate and coating material both adhered more strongly to a third substance (the contaminant) than to each other, or if some chemical bonds were thereby established.

Experimentally, the realization of an uncontaminated surface is limited to specific instances, such as that of a crystal which has been cleaved in vacuum immediately prior to evaporation, or the production of the substrate itself by an evaporation process. Of some interest, therefore, is the behavior of films on more commonly prepared surfaces.

CONDITIONS OF EVAPORATION

In the present work a well-baffled system was used with a dry ice-acetone trap, and the average pressures prevailing at the time of evaporation ranged from 10^{-4} to 10^{-5} mm Hg. Evaporation took place such that the film was deposited at a rate of the order of 500A in thickness per second at a point on the axis of symmetry of the system.

Two substrates were located symmetrically with respect to the axis of the system: one was a previously weighed flake of mica about 2 cm square, the other a glass microscopic slide. The former substrate was used for the measurement of the film's mass per unit area and its state of cleanliness was unimportant under the assumption that all evaporated atoms striking it remained there. The film on the glass slide was used for an interferometric measurement of thickness, and had been cleaned chemically and scrubbed and dried on a grease-free paper towel. Beyond this effort no pretense was made that the slides were actually uncontaminated.

MEASUREMENTS OF FILM DENSITY

The masses were determined with a semimicro balance with an error of ± 0.01 mg or less, which, since the masses involved were from 0.25–2.00 mg, corresponded to weighing errors of from 0.5 percent–4.0 percent. In order to determine the mass per unit area of these irregularly shaped films, an image was produced on photographic paper by a contact process, and then measured with a polar planimeter. In the measurements of copper density a colorimetric determination of mass was used in addition to the direct weighing and agreed in each case to within the experimental error.

Film thickness was determined with a multiple beam interferometer of the Tolansky type.[6] The deposited film was wiped away in regions to produce a series of parallel channels. A second coating of silver was evaporated onto this, which thus had steps in it equal to the channel depth, i.e. thickness of the underlying film. An interference pattern was next produced with the fringes crossing these channels and the channel depth was measured directly in terms of the shift of the 5461A mercury green line. The film thickness could be directly read to 50A and a series of such measurements reduced the probable error to half this, or better.

The results of these measurements for films of silver and copper are shown in Figs. (1) and (2).

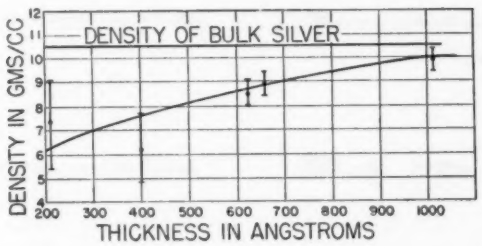


FIG. 1. The apparent density of a thin film of silver as a function of thickness.

DISCUSSION

To understand the meaning of these results we may consider several stages (Fig. 3) in the growth of two films, one of which is formed on a substrate providing tight binding (which generally appears to mean "clean"), the other providing loose binding (contaminated). After both films have reached an average thickness t (this being the quantity measured with the interferometer), it is clear that the departure from the expected density arises from migration effects at the interface.

The model shown in Fig. (3) is also supported by other evidence. Electron micrographs of thin films show light and dark regions which can be interpreted as due to different electron path lengths in the film. Perhaps less ambiguous are the results obtained in this laboratory on the polarization capacity of evaporated films based on an experiment of Bowden and Rideal[7] and modified by Wagner.[8] The actual area of a rough

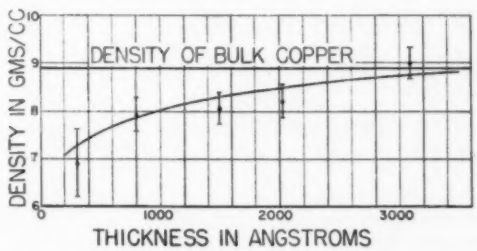


FIG. 2. The apparent density of a thin film of copper as a function of thickness.

[6] S. Tolansky, *Multiple Beam Interferometry* (Oxford University Press, New York, 1948).

[7] F. P. Bowden and E. V. Rideal, Proc. Roy. Soc. (London) **A120**, 59, 80 (1928).

[8] C. Wagner, J. Electrochem. Soc. **97**, 71 (1950).

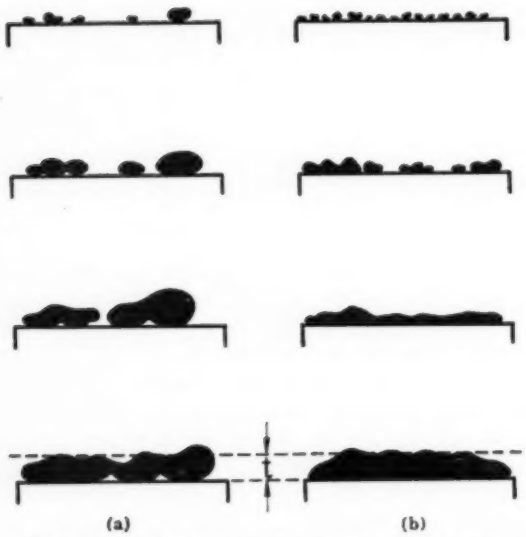


FIG. 3. Hypothetical stages in film growth. (a) Migration effect pronounced. (b) Little migration taking place. (Both films of same average thickness.)

metallic surface can be determined by measuring the transient current in an electrolytic cell, in which the surface is connected as the cathode, such current ceasing when the cell becomes polarized because of complete coating of the cathode by hydrogen atoms. A mercury surface is assumed to have a ratio of actual to apparent area of unity and is frequently used as a standard.

Silver films deposited on glass previously cleaned as described above were found by this method to have a ratio of real to apparent area of approximately 4 or 5. In these measurements, the metal films floated free from the substrate when placed in the electrolyte and the ratio quoted refers to the average of the two sides of the film. From considerations of surface energy and experimental knowledge concerning the outer, metal-air surfaces, it is fairly certain that this latter surface is reasonably smooth. The interfacial surface is therefore probably even rougher than the ratio above indicates because of the diluting effect of the outer surface in the average which is the observed quantity. The actual roughness is probably even more pronounced than indicated in Fig. (3).

Interface roughness can be reduced or eliminated if sufficient care is exercised and an appropriate substrate chosen. Films of finite conductivity only a few atom layers thick have been prepared[9] by using very high vacuums, well-cleaned and bombarded surfaces, and depositing on a low temperature substrate. In other instances, such as the use of organic substrates, these rigorous measures cannot all be taken and considerable deviations from a smooth surface must be accepted.

In recent experiments on the reflection of x-rays from evaporated films deviations from bulk density were evident.[10] For one particular copper film, measurements on reflectivity as a function of angle of reflection for x-rays indicated that for agreement with theory a density 11 percent below bulk density would have to be assumed. The film was dissolved from the mirror and the density measured by a colorimetric method. It was found to be 89 percent of the bulk value. This value appears in Fig. (2) at 1500A thickness. There is thus much evidence from different types of experiments for the reduced apparent density of evaporated films.

There is, however, no evidence that if one considers a sufficiently small volume of film, which still contains a great many atoms, he will find a density different from that of a bulk sample of equal purity. What one observes here then is not actually film density, but the nature of the film-substrate interface. However, because the wide use of evaporated films in experimental physics as targets, scattering foils, sources, surfaces, etc., it is of some importance that detailed assumptions regarding film structure take into account this interface porosity, and that bulk density values be used with care.

[9] Reinders and Hamburger, Ann. Physik. **10**, 649, 790, 905 (1931).
[10] L. M. Rieser, Jr., dissertation, Stanford University (1952).

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 3 MARCH, 1954

Noise in Semiconductors: Spectrum of a Two-Parameter Random Signal

STEFAN MACHLUP*
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received July 15, 1953)

With intended application to the theory of current noise in semiconductors, the spectrum is calculated of a random signal which may be in one of two states (YES-NO: random telegraph signal), where the mean lives σ and τ of the two states may be different. The form of the spectrum is the same as for the case with equal lives, the single parameter τ for that (classical) case now being replaced by $2[(1/\sigma)+(1/\tau)]^{-1}$; i.e., the rates of transition are additive.

1. NOISE MECHANISM

SEVERAL mechanisms proposed to explain current-dependent noise in semiconductor structures involve the idea that carriers may be caught in traps, thus remaining, for the time they are held captured, unavailable for conduction. Such a process would be at least to some extent random, and thus be responsible for statistical fluctuations in current, i.e., noise.

We may focus our attention either on a carrier or on a trap. In the first case, we have the possibilities (1) carrier captured, (2) carrier free. In the second case, (1) trap filled, (2) trap empty. The statistics of a single carrier will suffice for the calculation of the noise spectrum if the carriers are independent (uncorrelated); i.e., the probability of trapping a carrier does not depend on whether other carriers are captured or free. Trapping (and releasing) of a single carrier is then a purely random process. It follows that the noise powers due to the separate carriers are additive.

The origin of possible correlation between carriers is the exclusion principle, which limits the number of carriers in one trap to either zero or one. The probability of trapping is therefore proportional to the number of empty (available) traps. A sufficient condition for the absence of correlation is that a large fractional change in the number of empty traps be very improbable. This will be satisfied if the average number of (identical) empty traps is much greater than unity (Law of Large Numbers).

On the other hand, we may concentrate on the traps. If these are identical, we have a completely analogous situation: the statistics of a single trap suffice if the traps are uncorrelated—i.e., the probability of filling a trap does not depend on whether other traps are full or empty. Filling (and emptying) of a single trap is then a purely random process. A sufficient condition for this situation (traps independent) is that a large fractional change in the number of carriers be very improbable. This condition will be satisfied if the average number of free carriers is a large number.†

The two cases—carriers independent, traps independent—are by no means mutually exclusive. If both of the averages (number of empty traps, number of free carriers) are large numbers, both points of view lead to purely random processes.

2. SPECTRUM

In either of these favorable cases, we are dealing with a purely random signal which may be in one of two states, henceforth called 1 and 0; i.e., a random "telegraph signal." The probability of making a transition to 0 in a short time dt, given that we are in 1, is proportional to dt, say dt/σ. Similarly, the probability of making a transition to 1 in a time dt, given that we are in 0, is dt/τ; i.e., $1/\sigma$ and $1/\tau$ are the decay rates of the conditional occupation probabilities of the states 1 and 0, respectively. It follows that σ and τ are the mean lifetimes of the respective states. If we ask for the probability of being, at any time, in state 1, it is evidently $\sigma/(\sigma+\tau)$; the probability of being in state 0 is $\tau/(\sigma+\tau)$.

We shall begin by calculating the autocovariance of the signal, and shall then apply the Wiener-Khinchin formula to obtain the spectrum. This formula states that the spectrum is simply the Fourier transform of the autocovariance. Let the signal be $x(t)$; x may take on the values 0 and 1. All statistical properties of x are independent of the origin of t (stationary stochastic process).

The autocovariance is

$$A(s)\equiv\langle x(t)x(t+s)\rangle_{\text{average}}$$
$$=\sum_{ij} x_i x_j\cdot\{\text{Prob. that } x(t)=x_i\}\cdot\{\text{Prob. that } x(t+s)=x_j \text{ given that } x(t)=x_i\}. \quad (1)$$

Our fortunate choice of the state name "0" makes all but one term of this (four-term) sum vanish; i.e.,

$$A(s)=1\cdot 1\cdot\{\text{Prob. that } x(t)=1\}\cdot\{\text{Prob. of an even number of transitions in time } s, \text{ starting in state 1}\} \quad (2)$$
$$=\frac{\sigma}{\sigma+\tau}P_{11}(s).$$

* Present address: University of Illinois, Urbana, Illinois.

† If these averages are much less than unity, we again have statistical independence. Only if both of them are of the order of unity do we have appreciable correlations—and a knotty statistical problem.

Writing, similarly,

$P_{10}(s)=\{$Prob. of an odd number of transitions in time s, starting in state 1$\}$

we have

$$P_{11}(s)+P_{10}(s)=1. \tag{3}$$

We derive a differential equation for P_{11} as follows: For a small increment ds, we have

$$P_{11}(s+ds)=P_{10}(s)\frac{ds}{\tau}+P_{11}(s)\left(1-\frac{ds}{\sigma}\right). \tag{4}$$

In words: the probability $P_{11}(s+ds)$ of an even number of transitions in $s+ds$, starting in state 1, is the sum of the probabilities of two mutually exclusive events: (1) an odd number of transitions in s and one transition in ds, (2) an even number of transitions in s and no transition in ds. The increment ds is supposed small enough that the probability of *more* than one transition during it is negligible. Substituting from (3) we eliminate P_{10}; passing to the limit $ds\to 0$ we obtain the differential equation for $P_{11}(s)$,

$$\frac{dP_{11}}{ds}+\left(\frac{1}{\sigma}+\frac{1}{\tau}\right)P_{11}=\frac{1}{\tau}. \tag{5}$$

Evidently $P_{11}(0)=1$, i.e., in zero time we can have only zero transitions. With this initial condition we obtain the solution

$$P_{11}(s)=\frac{1}{\sigma+\tau}\left[\tau\exp\left\{-\left(\frac{1}{\sigma}+\frac{1}{\tau}\right)s\right\}+\sigma\right]. \tag{6}$$

We then substitute in (2) to obtain the autocovariance:

$$A(s)=\frac{\sigma}{(\sigma+\tau)^2}\left[\tau\exp\left\{-\left(\frac{1}{\sigma}+\frac{1}{\tau}\right)s\right\}+\sigma\right]. \tag{7}$$

So far we have implicitly limited outselves to positive times s. But we may just as well ask for the probability that an event "has occurred" in the past. Evidently all probabilities here are symmetric in past and future; hence

$$A(s)=A(-s), \tag{8}$$

and (7) holds for all s if we replace s by $|s|$ on the right-hand side.

We now take the Fourier transform of the autocovariance to obtain the spectrum (Wiener-Khinchin formula):

$$S(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}A(s)\exp(-i\omega s)ds$$

$$=\frac{1}{\pi}\frac{\sigma\tau}{(\sigma+\tau)^2}\frac{\left(\frac{1}{\sigma}+\frac{1}{\tau}\right)}{\omega^2+\left(\frac{1}{\sigma}+\frac{1}{\tau}\right)^2}+\left(\frac{\sigma}{\sigma+\tau}\right)^2\delta(\omega). \tag{9}$$

The δ function (dc component) comes from Fourier-transforming the constant term and evidently is not "noise." We display it here in order to preserve the normalization relation

$$\int_{-\infty}^{\infty}S(\omega)d\omega=\langle x^2\rangle=\frac{\sigma}{\sigma+\tau} \tag{10}$$

(conservation of power).

The simplicity of (9) is seen by writing it using the time constant T,

$$1/T=(1/\sigma)+(1/\tau), \tag{11}$$

$$S(\omega)=\frac{1}{\pi}\frac{\sigma\tau}{(\sigma+\tau)^2}\frac{1/T}{\omega^2+(1/T)^2}+\text{dc term.} \tag{9a}$$

A spectrum proportional to $\mu/(\omega^2+\mu^2)$, where μ is some "rate" constant, is a property of linear Markoff processes, of which the two-time-constant random telegraph signal is a special case. It is interesting to note that the rate $\mu=1/T$ is the sum of the rates $1/\sigma$ and $1/\tau$ of two correlated single-time-constant Markoff processes ($0\to 1$ transitions and $1\to 0$ transitions). We see also that the spectrum (except for the dc term) is symmetric in the two time-constants, i.e., independent of the state names.

3. LIMITING CASES

Equation (9) will reduce to well-known formulas in two cases.

A. Single Time Constant

If $\sigma=\tau$, i.e., the two states have the same mean lives, we have the case treated by Rice,[1] obtaining

$$S(\omega)=\frac{1}{\pi}\frac{1}{4}\frac{2/\tau}{\omega^2+(2/\tau)^2}+\text{dc term.} \tag{12}$$

B. Shot Noise

Suppose one of the time constants is very short; say $\sigma\ll\tau$: State 1 occurs but a small fraction of the time and has little probability of lasting long. In the limit we can treat double transitions $0\to 1\to 0$ as independent events. More precisely, the probability of a $0\to 1$ transition being followed a given time later by another $0\to 1$ transition is independent of the life of the intervening 1 state. This is similar to the type of model one postulates for shot noise; the difference is that in the shot noise case the individual pulses are assumed identical, while here the life of the 1 state, while likely to be short, is still a random variable.

Our formula (9) becomes

$$S(\omega)_{\sigma\ll\tau}=\frac{1}{\pi}\frac{\sigma}{\tau}\frac{1/\sigma}{\omega^2+(1/\sigma)^2}=\frac{1}{\pi}\frac{\sigma^2}{\tau}\frac{1}{1+\omega^2\sigma^2} \tag{13}$$

[1] S. O. Rice, Bell System Tech. J. **23**, 282 (1945), Sec. 2.7.

(omitting the dc term).[2] For angular frequencies ω low compared to $1/\sigma$, this gives a white spectrum:

$$S(\omega)\rightarrow\sigma^2/\pi\tau. \tag{14}$$

4. INTERPRETATION

Let us assign our symbols physical meaning relevant to noise in semiconductors. We identify the 1 state with the flow of a current $i=I$; i.e., a double-transition $0\rightarrow1\rightarrow0$ corresponds to a current pulse: a carrier leaves a trap and moves under the influence of an electric field until it is again trapped. The average rate of occurrence of pulses is $N=1/(\sigma+\tau)$. Changing to frequency notation ($f=\omega/2\pi$, only positive frequencies permitted), we write Eq. (9a) as mean square noise current per unit band width:

$$\begin{aligned}\langle i^2\rangle_f &\equiv I^2\cdot 4\pi S(2\pi f) \qquad (15)\\ &= 4I^2\frac{\sigma^2\tau^2}{(\sigma+\tau)^3}\frac{1}{1+4\pi^2f^2T^2}\\ &= 4I^2T^2/(\sigma+\tau) \text{ for low frequencies}\\ &= 4I^2T^2N. \qquad (15a)\end{aligned}$$

If n identical, uncorrelated carriers participate, their noise powers add, and we must multiply (15) by n. The average rate of occurrence of pulses is then $N=n/(\sigma+\tau)$, so that (15a) holds as it stands. In terms of the average dc current, $i_{dc}=nI\sigma/(\sigma+\tau)=I\sigma N$, the low-frequency form of (15) becomes

$$\langle i^2\rangle_f=\frac{4i_{dc}^2}{N}\frac{T^2}{\sigma^2}=\frac{4i_{dc}^2}{N}\left(\frac{\tau}{\sigma+\tau}\right)^2. \tag{16}$$

Equation (16) should be compared with the corresponding formula for pure shot noise (charge transferred per pulse is constant, $q=e$),

$$\langle i^2\rangle_{f\,\text{pure shot noise}}=2Ne^2=2i_{dc}e. \tag{17}$$

Pure shot noise is proportional to the first power of the average dc current; noise due to carrier trapping is proportional to its square.

The limiting case B of Section 3, $\sigma\ll\tau$, makes Eqs. (15) and (16) look more familiar, since in this approximation $T\cong\sigma$. The charge q transferred through the circuit in one pulse—the time integral of the current over the duration of the pulse—is I times the pulse length. As the mean pulse length is σ, the mean charge transferred per pulse is $\langle q\rangle=I\sigma$; the mean square charge per pulse (exponential distribution!) is $\langle q^2\rangle=2\langle q\rangle^2=2I^2\sigma^2$. We may write (15a) in terms of this latter quantity,

$$\langle i^2\rangle_{f(\sigma\ll\tau)}=4I^2\sigma^2N=2\langle q^2\rangle N; \tag{18}$$

again, it is instructive to compare this with the pure shot noise formula (17).

Eq. (16) for the mean square noise current as a function of average dc current becomes simply

$$\langle i^2\rangle_f=4i_{dc}^2/N. \tag{19}$$

We see that the noise given by (16) is always *less* than that given by the $\sigma\ll\tau$ approximation. Furthermore, the factor $T/\sigma=\tau/(\sigma+\tau)$ by which the general formula (16) is smaller than the approximation (19) has a straightforward interpretation in terms of the statistics of n independent carriers. Let n_f be the number of free (untrapped) carriers; then

$$\langle n_f\rangle/n=\sigma/(\sigma+\tau), \tag{20}$$

i.e., the average fraction of the carriers free at any instant is equal to the fraction of the time a single carrier is free. Therefore the factor we are interested in, $\tau/(\sigma+\tau)=1-\langle n_f\rangle/n$, is equal to the average fraction of the carriers trapped at any instant. We can say more: n_f has a binomial distribution since it is the result of n independent "choices" between two states. Hence its variance is

$$\begin{aligned}\langle(n_f-\langle n_f\rangle)^2\rangle &= n\left[\frac{\sigma}{\sigma+\tau}\right]\left[\left(1-\frac{\sigma}{\sigma+\tau}\right)\right]\\ &= n\frac{\sigma\tau}{(\sigma+\tau)^2}. \qquad (21)\end{aligned}$$

Dividing (21) by (20) we have for the ratio of the variance to the mean

$$\alpha=\frac{\langle(n_f-\langle n_f\rangle)^2\rangle}{\langle n_f\rangle}=\frac{\tau}{\sigma+\tau}, \tag{22}$$

which is again our interesting factor.

5. CONCLUSION

Treatments of noise due to carrier trapping[3] have tacitly assumed that trapping and "releasing" of carriers were uncorrelated processes. This is a good approximation if the mean "captured" time is long compared to the mean "free" time and leads to expressions of the form of our Eqs. (13), (18), and (19), in which only one time constant (the mean "free" time) determines the shape of the spectrum.[4] The general equations (9), (15), and (16), show how the spectrum depends on both time constants.

The author is indebted to H. C. Montgomery for many helpful discussions.

[2] This has the form of the expression obtained by J. Bernamont, Proc. Phys. Soc. (London) **49**, E138 (1937).

[3] A. Van der Ziel, J. Appl. Phys. **24**, 222 (1953); R. L. Petritz, Proc. Inst. Radio Engrs. **40**, 1440 (1952), also references cited therein.

[4] This approximation makes α (Eq. 21) $=1$; hence its use in Van der Ziel's (reference 3) "case (a)" does not seem justified.

Radiation Damage Effects on Order-Disorder in Nickel-Manganese Alloys*

LEWIS R. ARONIN
Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received July 23, 1953)

Effects of fast neutron irradiation in a nuclear reactor on order-disorder in a series of nickel-manganese alloys ranging from 16.5 to 31.9 atomic percent manganese have been studied by resistivity and magnetic induction measurements. Attainment of an "irradiated state" differing from either cold work or thermal disordering is suggested by comparison of exposure results on initially cold work disordered and initially thermally disordered Ni_3Mn. In a region from 16.5 to approximately 22 percent manganese, thermally disordered alloys are markedly affected by irradiation. These effects also appear to be superposed on the disordering produced by irradiation of initially ordered alloys.

From theoretical exponential dependence on period of irradiation, fast neutron disordering of Ni_3Mn is determined by relating the Bragg and Williams order parameter quadratically with resistivity and linearly with magnetic induction. This leads to an estimate of 5000 for the ratio of the number of atomic replacements to primary collisions with neutrons of energy in excess of 0.5 Mev.

I. INTRODUCTION

EFFECTS of nuclear radiations on the properties of solids have been studied by many investigators. Review articles have been given by Slater,[1] Dienes,[2,3] and Seitz.[4] Substitutional solid solution alloys which form ordered structures or "superlattices" are of value for fundamental studies in radiation damage. The disarrangement of atoms caused by fast particle irradiation permits estimates of atomic displacements even though most atoms may have returned to lattice sites.

From studies of electrical resistivity and magnetic induction, existence of an order-disorder transition about the Ni_3Mn alloy was first deduced by Kaya and Kussmann[5] in 1931. This transition is well suited to irradiation study since electrical resistivity decreases markedly and the alloy becomes strongly ferromagnetic on passing from the disordered to the ordered state. In the investigations reported herein, the effects of fast neutron irradiation in a nuclear reactor have been studied by resistivity and magnetic induction measurements on a series of nickel-manganese alloys ranging from 16.5 to 31.9 atomic percent manganese. A comparison of irradiation effects on both initially ordered and initially disordered alloys has been obtained.

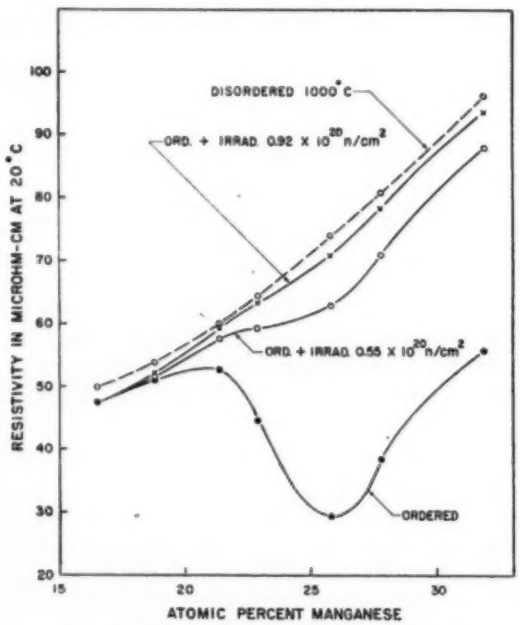

DISORDERED 1000°C
ORD. + IRRAD. 0.92 X 10^20 n/cm^2
ORD. + IRRAD. 0.55 X 10^20 n/cm^2
ORDERED
RESISTIVITY IN MICROHM-CM AT 20°C
ATOMIC PERCENT MANGANESE

FIG. 1. Effect of irradiation on resistivity of initially ordered alloys.

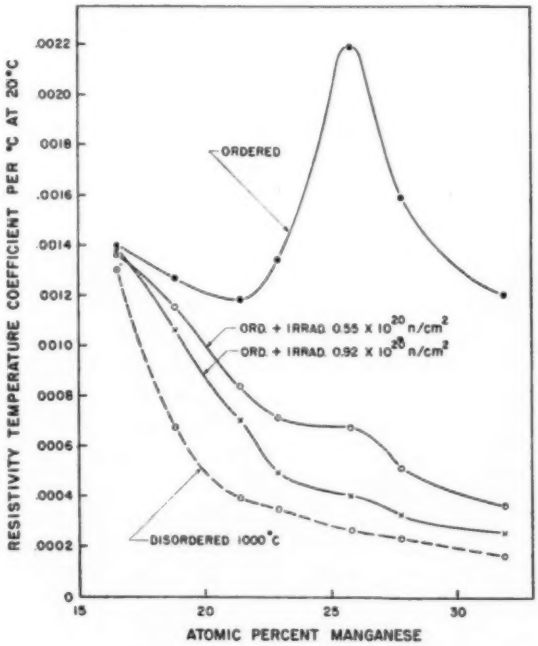

ORDERED
ORD. + IRRAD. 0.55 X 10^20 n/cm^2
ORD. + IRRAD. 0.92 X 10^20 n/cm^2
DISORDERED 1000°C
RESISTIVITY TEMPERATURE COEFFICIENT PER °C AT 20°C
ATOMIC PERCENT MANGANESE

FIG. 2. Effect of irradiation on resistivity temperature coefficient of initially ordered alloys.

* Work performed at the M.I.T. Metallurgical Project under U. S. Atomic Energy Commission Contract AT(30-1)-981.

[1] J. C. Slater, J. Appl. Phys. 22, 237 (1951).
[2] G. J. Dienes, Ann. Rev. of Nuclear Sci. II, 187 (1953).
[3] G. J. Dienes, J. Appl. Phys. 24, 666 (1953).
[4] F. Seitz, Phys. Today 5, No. 6, 6 (1952).
[5] S. Kaya and A. Kussmann, Z. Physik 72, 293 (1931).

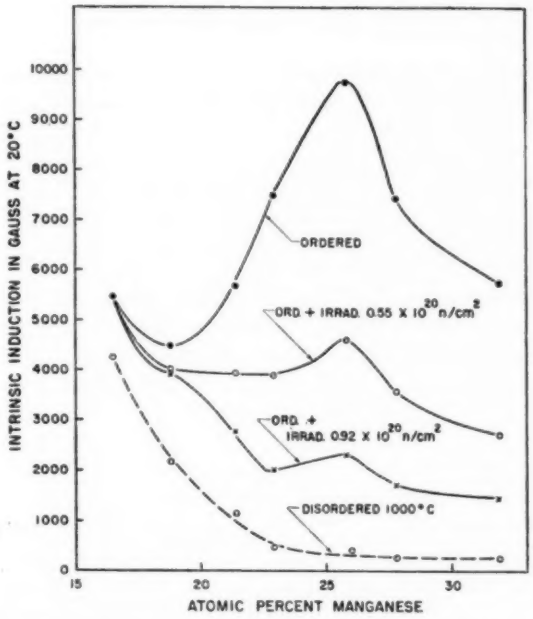

INTRINSIC INDUCTION IN GAUSS AT 20°C
ATOMIC PERCENT MANGANESE
ORDERED
ORD + IRRAD. 0.55 X 10^20 n/cm^2
ORD. + IRRAD. 0.92 X 10^20 n/cm^2
DISORDERED 1000°C

FIG. 3. Effect of irradiation on intrinsic induction of initially ordered alloys. Applied field = 20 000 oersteds.

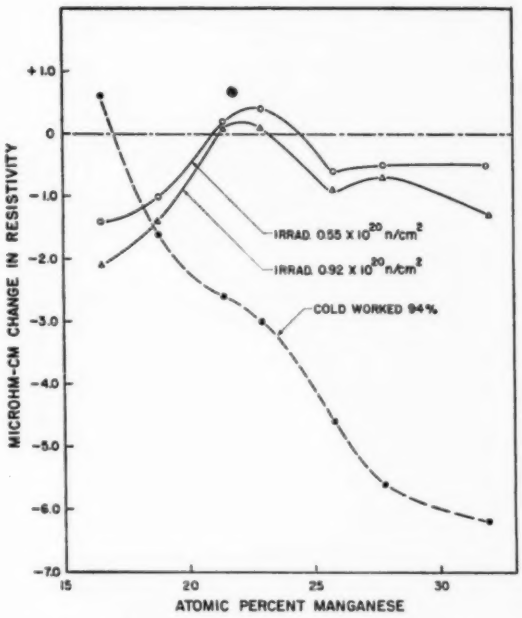


FIG. 4. Effect of irradiation on resistivity of initially thermally disordered alloys.

II. EXPERIMENTAL

Wire specimens $\frac{1}{16}$ in. in diameter by $4\frac{1}{2}$ in. long were fabricated from melts prepared by open-pot induction melting of electrolytic manganese and cobalt-free Mond nickel. Materials were melted under an argon blanket and each melt was cast into a hot-topped graphite mold to minimize pipe and segregation. Following a large amount of hot rolling at 1070°C, ingots were cold swaged to $\frac{1}{4}$-in. diameter with appropriate annealing between passes. A final 1000°C anneal and quench was performed, after which the material was cold swaged through a reduction of 94 percent to $\frac{1}{16}$-in. diameter wire.

Order-disorder heat treatments were performed with specimens sealed in evacuated quartz tubes. Specimens examined in this investigation correspond to the following three initial conditions:

(1) As disordered by 94 percent cold working from the annealed and quenched condition.

(2) Disordered by cold working 94 percent followed by annealing at 1000°C for one hour and water quenching.

(3) Ordered from condition (2) by slow cooling from 600°C to 300°C over a period of 10 days. The average cooling rates used were essentially 20°C per hour from 600°C to 500°C and 20°C per day from 500°C to 300°C.

Resistivity measurements were obtained by the current and potential method in the temperature range 20° to 50°C. Values of intrinsic induction ($4\pi I$) at 20°C in applied fields up to 20 000 oersteds were determined by ballistic galvanometer methods.

Fast neutron exposures for two different periods were obtained with specimens located in the nuclear reactor at a position where it is believed that specimen temperatures did not exceed 50°C during irradiation. The integrated fast flux exposures to neutrons having energy in excess of 0.5 Mev are estimated to be 0.55×10^{20} neutrons/cm² and 0.92×10^{20} neutrons/cm², respectively, for the two periods of irradiation.

III. RESULTS ON INITIALLY DISORDERED ALLOYS

Properties of initially thermally disordered alloys are indicated in the dashed curves of Figs. 1–3. The magnitude of property change induced by irradiation of these alloys may be seen from the solid curves of Figs. 4–6. Measurable alteration of parameters will be noted, with effects not linearly proportional to period of exposure. In each instance effects are most pronounced in the region from 16.5 to approximately 22 percent manganese.

For comparison with effects of irradiation, the dashed curves of Figs. 4–6 show the results of 94 percent cold work on initially thermally disordered alloys. Over most of the composition range the resistivity of cold worked alloys is lower than for thermal disordering (annealed and quenched) only. At the Ni_3Mn composition the decrease is over 5 percent. Magnetic induction values for disordered alloys decrease from ferromagnetic toward paramagnetic behavior. Cold worked alloys have lower magnetic values than quenched alloys.

Inasmuch as ordering results in greatly reduced resistivity accompanied by marked increase in magnetic induction, it would appear unlikely that the cold worked and quenched conditions represent only slight differences in residual order. It might be postulated that

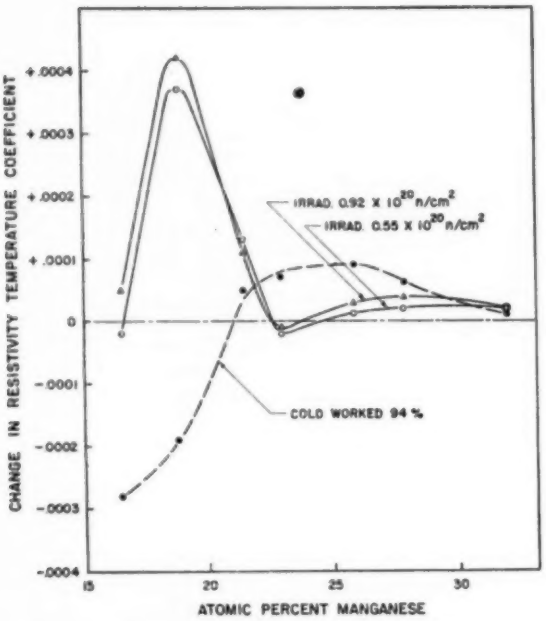


FIG. 5. Effect of irradiation on resistivity temperature coefficient of initially disordered alloys.

quenching leaves small nuclei of order which act as scattering centers for conduction electrons and at the same time act as small regions of high permeability. If these small nuclei were destroyed by subsequent cold working, one might account for both the decrease in resistivity and magnetic induction. On the other hand, this postulate requires considerable manipulation to be consistent with the effects of irradiation.

It has also been suggested that the decrease in resistivity could be caused by an increase in the Debye characteristic temperature brought about by cold work. The decrease in magnetic induction could be accounted for on the basis of the profound effect of residual strain on magnetic properties.

Of the properties measured, electrical resistivity of thermally disordered alloys (Fig. 4) exhibits the least sensitivity to irradiation. Near the Ni_3Mn composition, the effects are smallest—the curves for irradiated material actually intersecting the resistivity values for unirradiated material at two points between 21 and 25 percent manganese. On either side of this region, resistivity is decreased by irradiation, the maximum observed effect being a 4 percent decrease at 16.5 percent manganese. The effect on resistivity temperature coefficient (Fig. 5) is somewhat more striking—large increases being observed between 16.5 and 22 percent manganese.

Magnetic induction also exhibits pronounced shifts due to irradiation as may be seen from Fig. 6. From 16.5 to approximately 22 percent manganese, irradiation raises appreciably the induction values for thermally disordered alloys. Above 23 percent manganese the effect is reversed and the weak magnetic behavior of alloys in this region is further decreased by irradiation.

A simple cold work interpretation of radiation damage would appear to be precluded by consideration of the above data since:

(1) this would require that magnetic induction values of thermally disordered alloys be lowered by irradiation throughout the alloy range;

(2) the resistivity temperature coefficient for the lower manganese contents is shifted further by irradiation from values for cold worked material;

(3) irradiation does not alter resistivity of thermally disordered alloys consistently in the direction of cold work.

Figure 7 gives a comparison of irradiation effects on the alloy nearest the Ni_3Mn composition (the 25.8 percent Mn alloy) for both the thermally disordered and 94 percent cold work disordered conditions. Although electrical resistivity of the initially thermally disordered alloy is only slightly decreased by irradiation, the resistivity of the cold worked material is raised over 5 percent to a value approximately the same as for thermal disorder. It may be seen also that resistivity temperature coefficient and magnetic induction are changed to approximately the same values for material irradiated from initially cold worked and initially quenched conditions. These results suggest the attainment of an "irradiated state" which may be independent of the initial conditions and differs from either cold work or thermal disordering. Additional exposures for longer periods would be desirable to check conclusively whether an identical saturation limit is being approached.

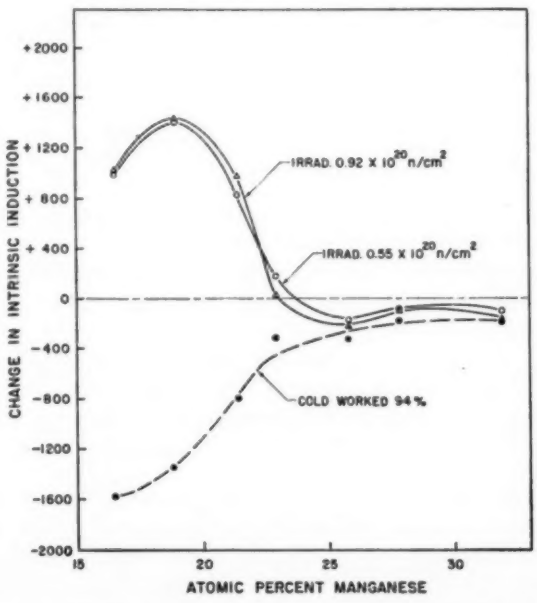


FIG. 6. Effect of irradiation on intrinsic induction of initially thermally disordered alloys. Applied field = 20 000 oersteds.

IV. RESULTS ON INITIALLY ORDERED ALLOYS—QUANTITATIVE EVALUATION

In Figs. 1–3 and Fig. 8 are indicated the effects of irradiation on the behavior of ordered alloys. At the lower manganese contents it is evident that some process other than simple disordering is also operative since resistivity values are lower and resistivity temperature coefficient and magnetic induction values are higher than would otherwise be anticipated for irradiated material. By comparison, it would appear that the previously discussed radiation effects as measured on initially quenched alloys may be regarded as also being approximately superposed on the disordering produced by irradiation of initially ordered material. Above 23 percent manganese, however, the principal effect is disordering, and the analysis given below does not require an understanding of the nature of what has been called the "irradiated state."

Desired results of irradiation study of an ordered alloy are an estimate of the change in degree of order with exposure and an evaluation of the number of atomic sites undergoing replacement per primary fast neutron collision. The precision of such estimates depends on the following considerations:

(1) the relationship between resistivity and magnetic properties with degree of order in the alloy;

(2) functional dependence of degree of order on the fast neutron flux exposure;

(3) magnitude of the fast flux, its energy spectrum,

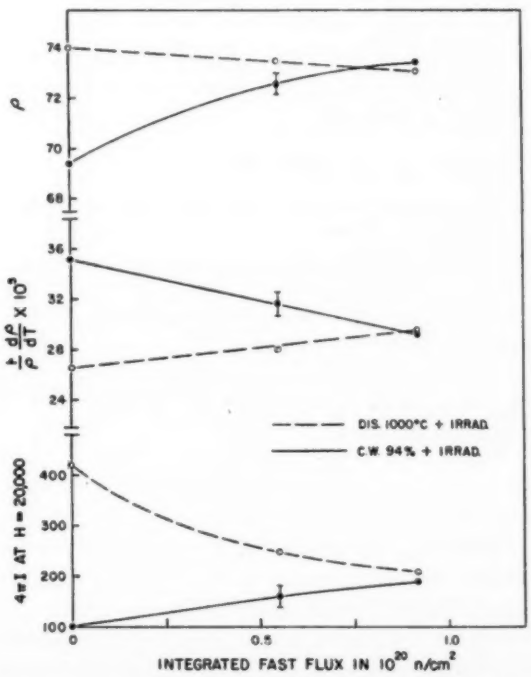


FIG. 7. Effect of irradiation on initially disordered Ni_3Mn.

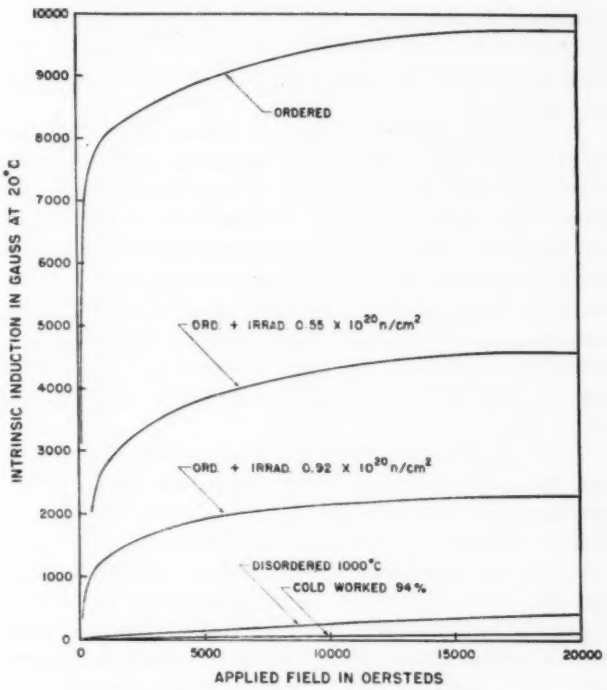

INTRINSIC INDUCTION IN GAUSS AT 20°C
ORDERED
ORD. + IRRAD. 0.55 X 10^{20} n/cm²
ORD. + IRRAD. 0.92 X 10^{20} n/cm²
DISORDERED 1000°C
COLD WORKED 94%
APPLIED FIELD IN OERSTEDS

FIG. 8. Intrinsic induction *vs* applied field—25.8 atomic percent Mn.

and damaging effect as a function of primary recoil energy as discussed by Seitz[6] and Ozeroff.[7]

The Bragg and Williams[8] definition of the long-range order parameter, S, is given as

$$S=\frac{p-r}{1-r} \quad \text{where} \quad 0\leqq S\leqq 1 \tag{1}$$

in which

r = fraction of sites occupied by B atoms in an alloy composed of A and B type atoms,

and

p = probability that a B site for the ordered alloy is occupied by a B atom.

From theoretical considerations, Muto[9] has given a quadratic dependence of resistivity on order parameter. A semiempirical solution of the following form has been given by Dienes:[10]

$$\rho=\rho_{\text{disorder}}-(\rho_{\text{disorder}}-\rho_{\text{order}})S^2, \tag{2}$$

where

ρ_{order} = resistivity at $S=1$

and

ρ_{disorder} = resistivity at $S=0$.

[6] F. Seitz, Discussions Faraday Soc., No. 5, 271 (1949).
[7] J. Ozeroff, Declassified U. S. Atomic Energy Commission Report, KAPL-205 (1949).
[8] W. L. Bragg and E. J. Williams, Proc. Roy. Soc. (London) **A145**, 699 (1934).
[9] T. Muto, Inst. Phys. Chem. Research Sci. Papers (Tokyo), **30**, 99 (1936).
[10] G. J. Dienes, J. Appl. Phys. **22**, 1020 (1951).

Hence, if resistivity values for the extreme conditions of order and disorder are known, the order parameter for any intermediate value of resistivity may be calculated.

It is recognized that Muto's theory might require some alterations in the case of transition elements such as Ni and Mn because of the change in magnetic moment with order. However, if resistivity is considered as resulting from incoherent scattering of electron waves by lattice periodicity imperfections, then Eq. (2) becomes analogous to S^2 dependence of the incoherent scattering of x-rays in an ordered lattice. Equation (2), therefore, could well have been assumed as a reasonable first approximation without reference to Muto's detailed theory.

From kinetic studies and data in the literature,[11,12] and by an examination of the irradiation data itself it appears that magnetic saturation may vary approximately linearly with order parameter, if order parameters are calculated on the basis of quadratic relationship with resistivity. This assumption does not appear to be in conflict with the concept that changes in order are accompanied by changes in nearest neighbor pairs with attendant variation in magnetic moment.[13,14]

The above correlation of resistivity and magnetic induction with the long-range order parameter may be subject to complexities introduced by order domain size effects.[15] However, if initial domain size is large, and disordering from irradiation is regarded as resulting in statistical rearrangement within domains without production of new sharp boundaries, error arising from this source will be of minor significance. For material irradiated in this investigation, it is believed that initial order as indicated from resistivity and magnetic parameters would be little influenced by further growth of domains.

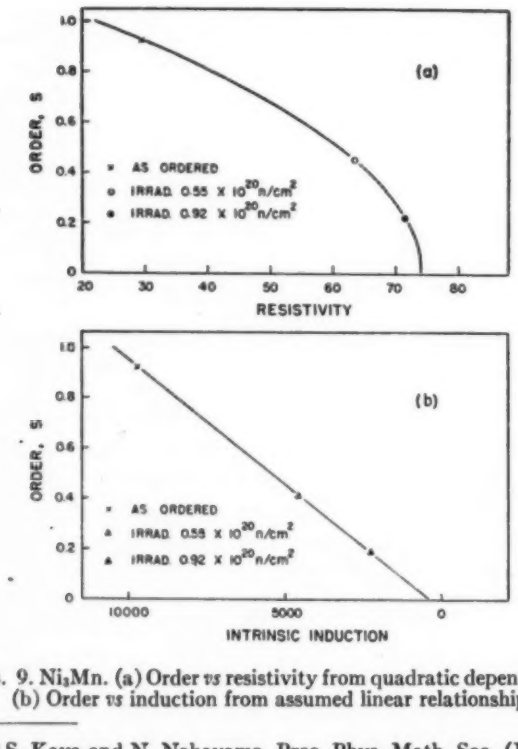


FIG. 9. Ni_3Mn. (a) Order *vs* resistivity from quadratic dependence. (b) Order *vs* induction from assumed linear relationship.

Exponential decrease in order parameter with fast flux exposure is anticipated on the basis of simple probability considerations. In these calculations it is assumed that duration of disturbance created by a primary collision is very short; that ordering forces in the region of disturbed atoms may, therefore, be neglected; and hence, these atoms assume random positions in the lattice.

At any instant the probability that a rightly occupied B site is disturbed and left disordered by irradiation in fast flux dn is given by $-k(1-r)pdn$, where k is a proportionality constant. The probability that a wrongly occupied B site is disturbed and left ordered by irradiation is $+k(1-p)rdn$. Hence, the total probability of disordering a B site becomes

$$dp=-k(1-r)pdn+k(1-p)rdn=-k(p-r)dn.$$

From Eq. (1),

$$p=(1-r)S+r \quad \text{and} \quad dp=(1-r)dS.$$

Therefore,

$$(1-r)dS=-k(1-r)Sdn,$$

which gives by integration,

$$S=S_0e^{-kn}. \tag{3}$$

On the average the total number of atomic replacements occurring on each lattice site due to exposure in fast flux dn is given by

$$dQ=kdn.$$

The number of primary collisions per atom in flux dn is

$$dC=[\sigma_A(1-r)+\sigma_Br]dn,$$

where σ_A and σ_B are fast neutron collision cross sections of A and B type atoms.

In the case of nickel-manganese $\sigma_{Mn}\cong\sigma_{Ni}$, and the number of atomic replacements per primary fast neutron collision becomes:

$$dQ/dC=k/\sigma. \tag{4}$$

Thus, if a value of k is obtained from Eq. (3) by relating measured properties after irradiation to order parameter, the number of atomic replacements per primary collision may be determined.

The above treatment would yield results capable of unique interpretation for irradiations in a known mono-

[11] S. Kaya and N. Nakayama, Proc. Phys. Math. Soc. (Japan) 22, 126 (1940).
[12] N. Volkenstein and A. Komar, Zhur. Eksper. i Teort. Fiz. 11, 723 (1941).
[13] J. E. Goldman, J. Appl. Phys. 20, 1131 (1949).
[14] R. Smoluchowski, J. phys. et radium 12, 389 (1951).
[15] L. R. Aronin, J. Appl. Phys. 23, 642 (1952).

energetic flux of fast neutrons. In this paper the calculation of the replacements per collision must be regarded for the present as an order of magnitude result.

For the fully ordered 25.8 percent Mn alloy, a minimum resistivity of about 22 microhm-cm is indicated from kinetic studies currently in progress. A resistivity of 74 microhm-cm is taken for complete disorder. By substitution of these values in the quadratic relationship, the order *versus* resistivity curve shown in Fig. 9(a) is obtained. An order parameter of 0.92 is indicated for the unirradiated ordered alloy. This is decreased by the two irradiation periods to 0.45 and 0.22, respectively.

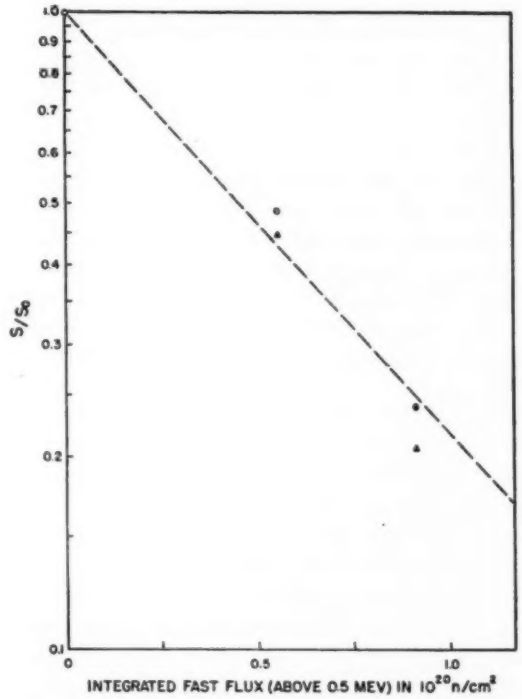


FIG. 10. $\mathrm{Ln}S/S_0$ *vs* integrated fast flux for Ni_3Mn.

In Fig. 9(b) is plotted an assumed linear relationship between order parameter and magnetic induction at a field of 20 000 oersteds. From Fig. 8 it may be seen that inductions at 20 000 oersteds are close to saturation values. The straight line of Fig. 9(b) is drawn between an induction value of 10 500 gauss for complete order (indicated by kinetic studies) and 400 gauss for complete disorder. An initial order parameter of 0.92 is also given by the value of induction (9750 gauss) for the ordered samples prior to irradiation. Induction values after the two irradiations then correspond to order parameters of 0.41 and 0.19, respectively, which are in reasonable agreement with values given above from resistivity data.

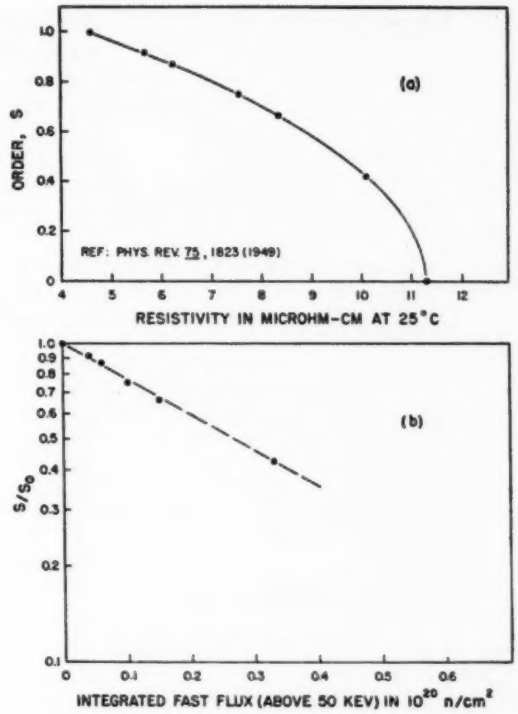


FIG. 11. Analysis of data of Siegel on Cu_3Au. (a) Order *vs* resistivity. (b) $\mathrm{Ln}S/S_0$ *vs* integrated fast flux.

In Fig. 10, a plot of $\ln S/S_0$ *versus* the integrated fast flux above 0.5 Mev is given. This leads to a value of $k \cong 1.5\times10^{-20}$ cm²/neutron in Eq. (3). If the fast neutron collision cross section for both nickel and manganese is taken to be 3×10^{-24} cm², the ratio of the number of atomic sites replaced to the number of primary collisions with fast neutrons having energy above 0.5 Mev is given by Eq. (4) as

$$\frac{dQ}{dC}=\frac{k}{\sigma}\cong\frac{1.5\times10^{-20}}{3\times10^{-24}}\cong 5000.$$

Siegel[16] has investigated the disordering of Cu_3Au by fast neutron bombardment. It is of interest to note that if the resistivity data obtained by Siegel is related quadratically to order parameter, good exponential dependence of order parameter on fast flux exposure is obtained as shown in Fig. 11. The number of atomic replacements per primary collision then given by Siegel's data appears to be comparable with results obtained above on Ni_3Mn.

ACKNOWLEDGMENT

The writer wishes to acknowledge helpful discussions with Professor A. R. Kaufmann, J. L. Klein, W. L. Lees, and W. B. Nowak of M.I.T., and Professor H. Brooks of Harvard University.

[16] S. Siegel, Phys. Rev. **75**, 1823 (1949).

Experimental Study of Low-Frequency Effects on the Dynamic Modulus of a Buna-N Rubber*

ALLEN Q. HUTTON, *Military Physics Research Laboratory, University of Texas, Austin, Texas*

AND

A. W. NOLLE, *Department of Physics, University of Texas, Austin, Texas*

(Received July 27, 1953)

The linear dynamic Young's modulus of a vulcanized Buna-N gum rubber is measured at frequencies of 0.05, 0.10, and 1.00 cps, in the temperature range −22°C to 30°C, by a method in which a small differential sinusoidal elongation is superimposed on a 5 percent static elongation. The width of the dispersion range on the temperature scale (the range in which the logarithm of the modulus increases steeply with decreasing temperature) is only about 10°C, as contrasted with widths as great as 25°C found in previous measurements on a similar compound at frequencies of several kilocycles. The modulus-temperature plot shifts upward by only about 4°C per decade of frequency increase in the present range, as contrasted with about 10°C per decade in the previous measurements at higher frequencies. It is concluded that this elastomer cannot be described properly by means of the "method of reduced variables," in which the dynamic properties are ascribed to mechanisms having identical temperature dependence, and that the low-temperature behavior is governed by mechanisms distinct from those effective in the audio-frequency range at room temperature.

INTRODUCTION

THE purpose of the work reported in the present paper is to extend certain previous measurements[1] on the dynamic mechanical properties of vulcanized Buna-N rubber. The relation of the present study to that previous work will first be explained briefly.

In a portion of the previous investigation,[1] the propagation constants for longitudinal waves in a thin strip of a Buna-N gum vulcanizate designated as B-1 (compound formulas are given at the conclusion of the present paper) were measured at various frequencies in the range 1000–20 000 cps, and values of the (complex) dynamic Young's modulus as a function of temperature and frequency were derived from the results. The sample was under 5 to 10 percent static elongation; hence, the result was designated as a differential dynamic modulus. It was found that the dispersion range on the temperature scale (that is, the range of temperature in which the logarithm of the real part of the modulus is increasing approximately uniformly with decreasing temperature) had a width of about 25C°, and that the center of the dispersion range moves upward in temperature by about 10C° for a 10-fold increase of frequency. These results were essentially confirmed at frequencies up to 70 kc, in the same investigation, by magnetostriction-rod measurements on a Buna-N gum vulcanizate designated B-100; in this case the width of the dispersion range and the effect of changing frequency were somewhat larger than the values given above. In the magnetostriction method, the experimental result was a differential modulus in the sense that the sample was under a static compression of several percent.

In another portion of the earlier investigation, the compound B-1 and also an unvulcanized Buna-N gum (Hycar OR-15) were examined by a vibrating-reed method. In this method, which depends upon observing the resonant frequency and the band width for a vibrating cantilever formed from the specimen, there is no static deformation of the sample, and, moreover, the frequency increases as the real part of the modulus increases. A typical frequency range for a single reed, as the temperature was varied, was 25 to 600 cps. The vibrating-reed results differed from the results cited for higher frequencies as follows: (a) the width of the dispersion range on the temperature scale was 15C° or less for the reed measurements as compared with a value of about 25C° for higher frequencies; (b) the dispersion range, as observed by the reed method, occurred at a higher temperature than that predicted from the longitudinal-wave results. The second point suggests that the temperature of maximum dispersion becomes less frequency-sensitive as the frequency is lowered. The same general results were found for a carbon-loaded Buna-N compound, B-5; point (b) was particularly pronounced in this case.

Because of the differences cited above between the vibrating-reed method and the methods used at higher frequencies, it was considered possible that the apparent differences (a) and (b) between results at low frequencies and those at high frequencies were introduced by the experimental techniques. To examine this possibility, it is desirable to make a set of low-frequency measurements by a method in which the sample is elongated differentially, in a fashion directly comparable to the local deformation in the longitudinal-wave experiment which was used at higher frequencies. Such low-frequency measurements are reported below. Since the frequency range of the present work is 0.05 to 1.0 cps, the lower frequency limit of the measurements is extended considerably below previous values.

* Based on a thesis submitted by Allen Q. Hutton in partial fulfillment of the requirements for the degree of Master of Arts in Physics at The University of Texas.

[1] A. W. Nolle, J. Polymer Sci. 5, 1–54 (1950).

DEFINITION OF TERMS

The differential dynamic Young's modulus reported in this study follows the convention used previously. The expression

$$E^* = E_1 + jE_2 \tag{1}$$

defines the complex modulus, where E_1 is a measure of the magnitude of elastic force which is in phase with a sinusoidal displacement, and E_2 is a measure of the viscous force which is out of phase with the displacement by $\pi/2$ radians. Both E_1 and E_2 have the dimensions dynes/cm^2.

For a sample of length l and cross-sectional area A, subjected to a pure elongation ψ, the force contribution F due to the real part E_1 of the Young's modulus is given by

$$F = E_1 A\psi/l. \tag{2}$$

The quantity E_2 may be defined by the expression $E_2 = \gamma\omega$, where γ is the "normal viscosity coefficient," which agrees with the geometrical definition of Young's modulus rather than with that of the usual shear viscosity coefficient, and ω is the frequency of applied stress in radians per second.

An additional relationship reported is the loss factor defined by the expression

$$E_1/E_2 = \cot\theta, \tag{3}$$

where θ is the phase angle of the mechanical impedance.

THE EXPERIMENTAL APPARATUS

An electromechanical low-frequency force generator is used to impart sinusoidal elongations to the sample. The magnitude of the elongation is detected by the effect of a change in capacitance on the frequency of an rf oscillator. A voltage proportional to the applied force is derived from the force generator and a voltage proportional to the magnitude of the differential elongation is supplied by a frequency discriminator circuit. The two voltages, when applied to the axes of a cathode-ray oscilloscope, produce a Lissajous figure display which provides a measure of the applied force, the resulting elongation of the sample, and the phase difference between the two. Figure 1 is a functional block diagram of the major components of the apparatus.

The sinusoidal force is derived from the interaction of a soft iron slug and the magnetic field of an ironclad solenoid which is energized with a sinusoidal current of the desired frequency. The low-frequency current source is comprised of a rotary inductor (Selsyn generator) driven mechanically by an electric motor through a variable speed transmission and a rectifier circuit to demodulate the generator output.

The displacement detector consists of a two-plate, constant-area, variable-separation condenser pickup in the tank circuit of an rf electron-coupled oscillator, and a frequency discriminator. One plate of the condenser pickup is affixed to the drive rod of the apparatus, and the other plate is mounted to the apparatus frame so that the variations in plate separation are equal to the magnitudes of the differential elongations of the sample.

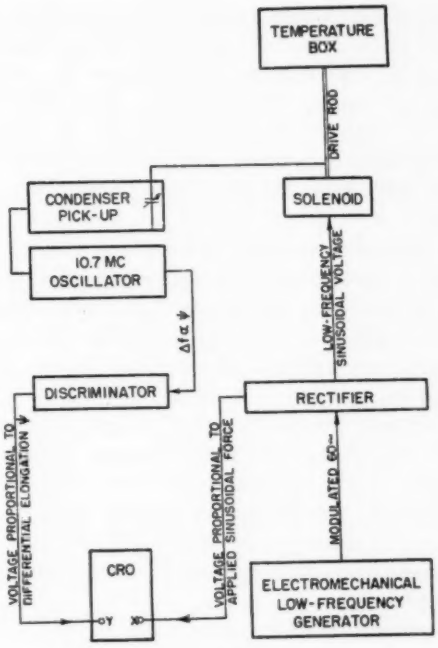


FIG. 1. Block diagram of experimental apparatus.

The rubber sample is in a thermostatically controlled box provided with forced air circulation. Solid carbon dioxide is used as the cooling agent, supplemented with coils attached to a constant-temperature bath.

Detection is by means of a cathode-ray oscillograph with long-persistence screen and high-gain dc amplifiers. The Lissajous figures displayed on the screen are photographed with a 35-mm oscillograph camera to furnish a film record of the fundamental data.

THE EXPERIMENTAL PROCEDURE

A strip of Buna-N rubber, compound B-100, of uniform rectangular cross section, 0.10 cm thick, 0.21 cm wide, and 9.74 cm in length, was used as the sample. Under the influence of the drive rod assembly, a 5 percent static stress was imposed on the sample. All calculations of modulus were based on the strained length and corresponding cross-section area of the sample.

The adjusted cross-section area was determined on the basis of the percent static elongation from the relationship

$$A = A_0/1.05, \tag{4}$$

where A is the cross-section area of the stretched sample, and A_0 is the cross-section area of the undistorted sample. This calculation assumes an unchanged volume of the sample for the deformation imposed.

All data were taken at selected temperature intervals from 30 to −22°C, the selection being based on the reaction of the sample to changes in temperature determined during a trial run in order that the steep portions of the modulus curves might be well defined. At the predetermined temperatures, data were taken for each of the three frequencies, 0.05 cps, 0.1 cps, and 1.0 cps.

The information obtained from the Lissajous figures provides a measure of the total mechanical impedance of the system from which the complex modulus of the rubber sample may be obtained when the mechanical resistance and mechanical reactance of the apparatus are known.

Equivalent mechanical circuits are derived for the various conditions of measurement, and the contributions of the apparatus to the total mechanical impedance of the oscillating system were determined by observing the response of the system when elements with known mechanical properties were substituted for the sample.

The general problem may be defined in the following way:

$$Z_m=\frac{F}{j\psi\omega}=R_m+jX_m, \tag{5}$$

where Z_m is the total mechanical impedance of the oscillating system, ψ is the displacement resulting from an applied force F, ω is the frequency in radians per second, R_m is the total mechanical resistance of the system, and X_m is the total mechanical reactance of the system.

The known contributions of the apparatus may be substituted into the expressions derived from the equivalent mechanical circuits and the values determined for the contribution R_r of the rubber sample to the effective mechanical resistance of the system and the contribution K_r of the sample to the mechanical stiffness of the system. Values for E_1 and E_2 may then be computed from the expressions

$$E_1=K_rl/A, \tag{6}$$

$$E_2=R_rl\omega/A. \tag{7}$$

The greatest source of error in the measurements arises in the value of deflection sensitivity used. The value was determined from an average of forty-eight different micrometer readings. While random variations about the average value chosen indicate a probable error of 3 percent in making the measurements of deflection sensitivity, dynamic measurements of the stiffness of a calibrated spring, using calibration data of the same nature, agreed with the spring constant with an error less than 1.5 percent.

In view of the existence of certain "end effects" at the clamps and an uncertainty in the true effective length and cross section of the sample, which cannot be precisely evaluated, the uncertainty in the actual value of modulus is considered to be somewhat greater than 2 percent.

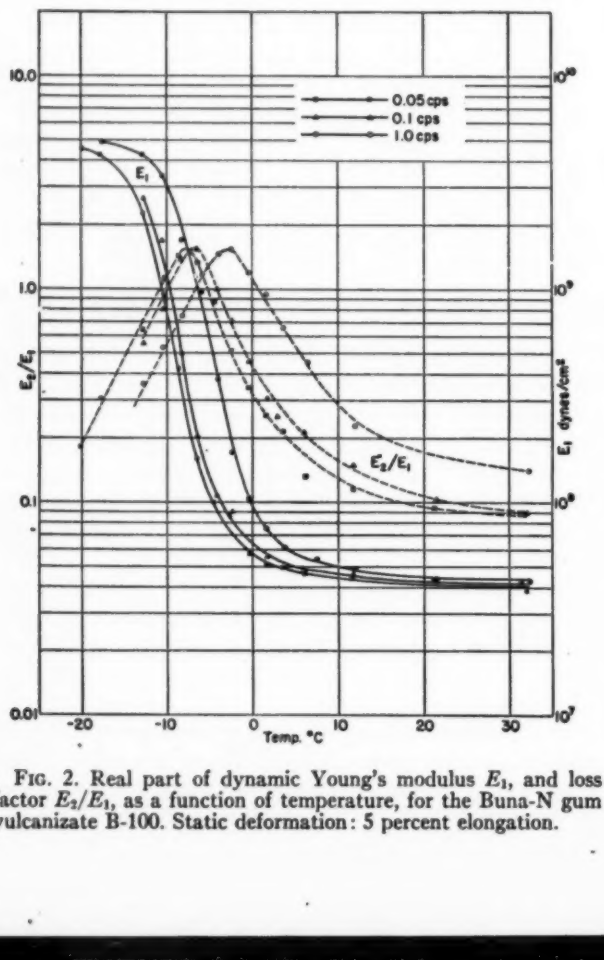


FIG. 2. Real part of dynamic Young's modulus E_1, and loss factor E_2/E_1, as a function of temperature, for the Buna-N gum vulcanizate B-100. Static deformation: 5 percent elongation.

RESULTS

Previous results in the low-frequency range[1] extrapolated according to certain conclusions drawn from other data show the following characteristics:

(1) A narrowing of the dispersion range in terms of temperature with decreasing frequency is indicated.

(2) Accompanying the narrowing of the dispersion range, the position of the dispersion range on the temperature scale becomes less frequency-sensitive as the frequency is reduced.

The results of the present investigation seem to confirm both conclusions. Figure 2 shows a plot of log E_1 and log E_2/E_1 *vs* temperature for each of the three frequencies of applied stress employed. A study of the results shows a dispersion range of the order of 10°C which appears to be consistent with the vibrating-reed results previously obtained at about 100 cps. The temperature of maximum dispersion does become less frequency-sensitive for frequencies of 1 cps and less. The results of the present study indicate a shift of the dispersion range toward lower temperatures for the decade from 0.1 cps to 1.0 cps of the order of 4°C. The previous results cited above show a shift of the dispersion range of the order of 10° per decade for frequencies of several kilocycles.

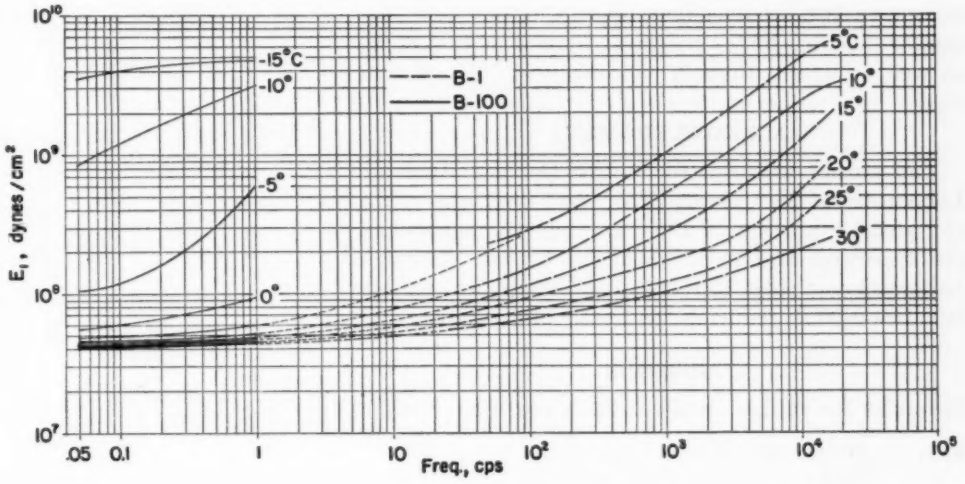


FIG. 3. Real part of dynamic Young's modulus E_1 at various temperatures as a function of frequency for the Buna-N gum vulcanizate B-100 of the present study together with results obtained with a similar Buna-N compound B-1 from previous measurements (see reference 1).

Note that the shift of the dispersion range is also evident in the position along the temperature scale of the plots of log E_2/E_1. The close correspondence of the loss peaks and the region wherein the real part of the modulus increases most rapidly is also apparent.

To assist in further interpretation of the results, the present data for the real part of the Young's modulus have been plotted in Fig. 3, together with values at higher frequencies as obtained with a similar compound in earlier measurements.[1] In the figure, the modulus as a function of frequency is shown for several different temperatures. The absence of a constant relation between the frequency derivative of the modulus (on logarithmic scales) and the derivative of logarithmic modulus with temperature is evident by inspection. For example, for temperatures in the neighborhood of 25C°, and frequencies in the range 100 to 1000 cps, the effect of a change of one decade in frequency can be compensated by a temperature change of some 12 centigrade degrees. On the other hand, for temperatures immediately above 0°C, and frequencies in the range 0.05 to 1.0 cps, a decade of frequency change corresponds to only about 3 centigrade degrees of temperature change.

Ferry[2] has pointed out that the "method of reduced variables" may be used to summarize the frequency-temperature behavior of polyisobutylene in a single plot of the shear modulus, and its loss factor, as a function of log (ωa_T), where ω is radian frequency and a_T is a function of temperature only. (A similar relation holds for the Young's modulus, which is closely three times the shear modulus.) The present work indicates, then, that the method of reduced variables does not apply to data for vulcanized Buna-N rubber over a wide range of temperature and frequency. That this would be the case was pointed out by Ferry, who showed that the method of reduced variables applies satisfactorily to data for vulcanized butyl rubber at frequencies in the kilocycle range and for temperatures above 0C°, but then remarked that the method could not be expected to apply to this compound at very low frequencies, where the cross-linked structure would govern the mechanical properties.

The conclusions from present results go further, however, than simply to confirm that the method of reduced variables is inapplicable at low frequencies to a cross-linked material. The values of modulus which are found in the lower part of our temperature range are much larger than those characteristic of kinetic-theory elasticity in a cross-linked structure, and apparently represent a phenomenon having no special relation to the existence of cross-linkage. The failure of the relationship between frequency effects and temperature effects to be the same at low frequencies as at high frequencies seems to argue that the mechanisms which govern the temperature dependence of the modulus at low frequencies may not be those viscoelastic mechanisms which are important for frequencies of the order of one kilocycle and higher. Thus, the concept of a low-temperature stiffening phenomenon, which, unlike the other effects, is a unique function of temperature, is not ruled out. The effect becomes observable at low frequencies only; when the frequency is sufficiently large, the low-temperature stiffening just mentioned is masked by the large increase of modulus with decreasing temperature which is characteristic of the viscoelastic mechanisms to which the reduced-variable concept does apply. This is a reassertion of the suggestions which were advanced on the basis of the earlier tentative low-frequency data.[1]

Recent low-frequency measurements of the dynamic shear modulus for several different elastomers are reported by Philippoff.[3] Unfortunately, a comparison of his results with those reported in this paper is limited, since his measurements thus far reported were all made at room temperature. It is interesting to note, however, that the modulus at very low frequencies

[2] Ferry, Fitzgerald, Johnson, and Grandine, J. Appl. Phys. 22, 712–722 (1951).

[3] Wladimir Philippoff, J. Appl. Phys. 24, 685–689 (1953).

becomes substantially independent of frequency in the case of the cross-linked polymers studied by Philippoff, as well as in the Buna-N vulcanizate of the present study.

ACKNOWLEDGMENT

The author is indebted to the Defense Research Laboratory of the University of Texas for the loan of a considerable portion of the electronic components of the apparatus and to the Military Physics Research Laboratory of the University of Texas for certain facilities used in reducing the data.

APPENDIX

Formulas for rubber compounds, in parts by weight:

B-1:	Hycar OR-15	100
	ZnO	5
	S	1.5
	Butyl 8 (a dithiocarbonate)	10 ml
	Methyl ethyl ketone	480
	20 minutes milling; solvent-cast	
B-5:	Hycar OR-15	100
	ZnO	5
	S	2
	Butyl 8	10 ml
	P33 black (soft thermal black)	50
	Methyl ethyl ketone	480
	Benzoic acid (gel inhibitor)	2
	15 minutes milling; solvent-cast	
B-100:	Hycar OR-15	100
	ZnO	5
	S	1.5
	Benzothiazyl disulfide	1.3
	60 minutes at 143°C	

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 3 MARCH, 1954

Estimation of Correlation Coefficients from Scatter Diagrams

GEORGE R. SUGAR
National Bureau of Standards, Washington 25, D. C.
(Received July 27, 1953)

Theoretical relations between statistical scatter diagrams of two random variables and their joint probability density functions are used to derive a simple method of estimating correlation coefficients. The cases considered are those of variables with Gaussian and Rayleigh distributions. Theoretical curves and experimental scatter diagrams are given for each case. Some factors which may influence the accuracy of the proposed method are discussed, and it is concluded that the errors in estimating correlation coefficient will be of the order ±0.1.

INTRODUCTION

THE use of scatter diagrams to examine the relationship between random variables is a well-known statistical practice.[1] By inspection of a scatter diagram the existence of correlation and the kind of correlation between two variables may be determined. However, the use of the scatter diagram to obtain a numerical estimate of the correlation coefficient is not well known and it is the purpose of this note to demonstrate how this may be accomplished in certain specific cases. A general method of procedure will be outlined first and then the two cases of variables with Gaussian and Rayleigh distributions analyzed. Examples of scatter diagrams obtained for each case will be given and a simple method for making the scatter diagrams by use of an oscilloscope discussed. Finally some of the factors which determine the accuracy of the proposed method of analysis will be discussed.

[1] P. G. Hoel, *Introduction to Mathematical Statistics* (John Wiley and Sons, Inc., New York, 1947), p. 78.

1. GENERAL PROCEDURE

Let x and y be two random variables each having the same probability density function, $f(x)$, $f(y)$, and a joint probability density function $g(x,y)$ containing a correlation parameter r. The curve $g(x,y)=k$, for suitably chosen values of the constant k, is the locus of constant joint probability density. However, in terms of a scatter diagram this is a locus of constant dot density (dots per unit area). Then, given a scatter diagram of the variables x *versus* y the loci of constant dot density may be sketched and then compared with the loci obtained from the equation $g(x,y)=k$. This gives an estimate of the true value of r for the scatter diagram. The correlation coefficient ρ_{xy} is defined by

$$\rho_{xy}=\frac{\langle xy\rangle-\langle x\rangle\langle y\rangle}{\sigma_x\sigma_y},$$

where $\langle x\rangle$ indicates the average value of x and $\sigma_x^2=\langle x^2\rangle-\langle x\rangle^2$. Since all of the quantities on the right of the equality can be obtained from $f(x)$, $f(y)$ and $g(x,y)$, it is possible to find the relation between ρ_{xy} and r.

2. ANALYSIS OF THE GAUSSIAN DISTRIBUTED CASE

The probability densities in this case are[2]

$$f(x)=\frac{1}{\sigma_x(2\pi)^{\frac{1}{2}}}\exp\left[-\frac{(x-\langle x\rangle)^2}{2\sigma_x^2}\right],$$

$$f(y)=\frac{1}{\sigma_y(2\pi)^{\frac{1}{2}}}\exp\left[-\frac{(y-\langle y\rangle)^2}{2\sigma_y^2}\right],$$

$$g(x,y)=\frac{1}{2\pi\sigma_x\sigma_y(1-\rho^2)^{\frac{1}{2}}}\exp\left\{-\frac{1}{2(1-\rho^2)}\right.$$
$$\times\left[\left(\frac{x-\langle x\rangle}{\sigma_x}\right)^2-2\rho\left(\frac{x-\langle x\rangle}{\sigma_x}\right)\left(\frac{y-\langle y\rangle}{\sigma_y}\right)\right.$$
$$\left.\left.+\left(\frac{y-\langle y\rangle}{\sigma_y}\right)^2\right]\right\},$$

where the correlation parameter is ρ, the correlation coefficient itself. For convenience assume $\langle x\rangle=\langle y\rangle=0$, $\sigma_x=\sigma_y=1$. Then the loci of constant joint probability density are given by

$$\frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}}\exp\left[-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)\right]=k$$

or

$$x^2-2\rho xy+y^2=-2(1-\rho^2)\ln[2\pi k(1-\rho^2)^{\frac{1}{2}}]=K(\rho,k).$$

For fixed values of ρ and k the locus obtained is an ellipse with its axes along the lines $y=x$, $y=-x$. The ratio R of the minor to major axes is related to the correlation coefficient by

$$\rho=\pm\frac{1-R^2}{1+R^2},$$

where the positive sign corresponds to the major axis in the first and third quadrants and the negative sign to the major axis in the second and fourth quadrants. The results in this case are especially simple since for a given ρ the ellipses for all k are of the same eccentricity. The ellipses for various values of ρ and k are shown in Fig. 1.

3. ANALYSIS OF THE RAYLEIGH DISTRIBUTED CASE

The probability densities in this case are[3]

$$f(x)=\frac{2x}{\langle x^2\rangle}\exp\left[-\frac{x^2}{\langle x^2\rangle}\right],\quad f(y)=\frac{2y}{\langle y^2\rangle}\exp\left[-\frac{y^2}{\langle y^2\rangle}\right],$$

$$g(x,y)=\frac{2xy}{\langle x^2\rangle-2r^2}I_0\left[\frac{2xyr}{\langle x^2\rangle-2r^2}\right]\exp\left[-\frac{(x^2+y^2)(\langle x^2\rangle)^{\frac{1}{2}}}{(\langle x^2\rangle-2r^2)\sqrt{2}}\right],$$

[2] P. G. Hoel, *Introduction to Mathematical Statistics* (John Wiley and Sons, Inc., New York, 1947), pp. 103 and 104.
[3] S. O. Rice, Bell System Tech. J. 24, 75 (1945).

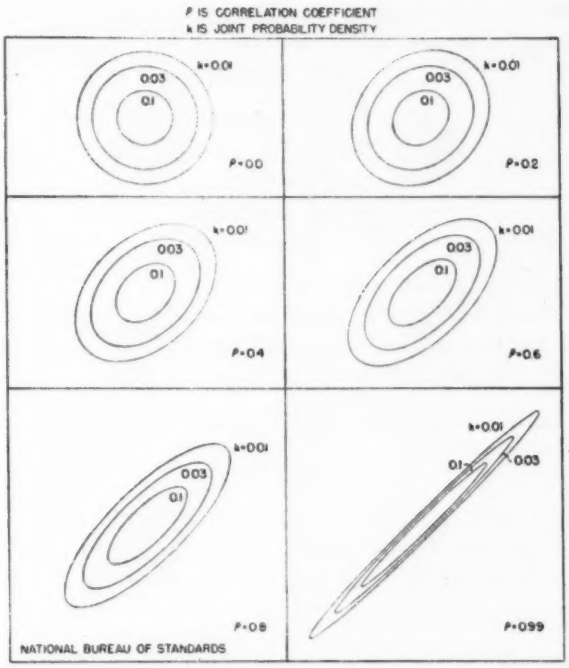


FIG. 1. Loci of constant joint probability density for normalized two-dimensional Gaussian distributions.

where $\langle x^2\rangle=\langle y^2\rangle$, r is a correlation parameter, and I_0 is the modified Bessel function of the first kind of zero order. For convenience set $\langle x^2\rangle=\langle y^2\rangle=2$. Then the desired loci are solutions of

$$\frac{xy}{1-r^2}I_0\left[\frac{xyr}{1-r^2}\right]\exp\left[-\frac{x^2+y^2}{2(1-r^2)}\right]=k.$$

In this case the loci are not easily determined by inspection of the equation. Moreover the form of the relation is not well suited to numerical computation. However with a suitable change of variables the relation becomes

$$\eta I_0\left[\eta\frac{r}{1-r^2}\right]=k(1-r^2)\exp\left[\frac{(\xi^2+\eta^2)^{\frac{1}{2}}}{1-r^2}\right].$$

If the left side is set equal to a constant the resulting equation is that of a straight line. Similarly the right side gives the equation of a circle. The problem is then easily solved (numerically) as the intersection of the line and circle. The transformation of variables used above,

$$x=(\xi^2+\eta^2)^{\frac{1}{2}}\left[1+\frac{\xi}{(\xi^2+\eta^2)^{\frac{1}{2}}}\right]^{\frac{1}{2}},$$

$$y=(\xi^2+\eta^2)^{\frac{1}{2}}\left[1-\frac{\xi}{(\xi^2+\eta^2)^{\frac{1}{2}}}\right]^{\frac{1}{2}},$$

then gives the required loci. The relation of ρ to r can be evaluated by use of the equations

$$\langle xy\rangle=\int_0^\infty\int_0^\infty xy\, g(x,y)\, dx\, dy,$$

$$\langle x\rangle=\langle y\rangle=\int_0^\infty x\, f(x)\, dx,$$

and the definition of ρ.

The first of these has been evaluated by Burgess[4] and the second may be integrated directly. The relation of ρ to r (as given by Bramley[4]) is

$$\rho=\frac{4[E(r)-\frac{1}{2}(1-r^2)K(r)]-\pi}{4-\pi}\approx r^2,$$

where $E(r)$ and $K(r)$ are complete elliptic integrals of the first and second kinds, respectively. The loci for various values of ρ and k have been computed and are shown in Fig. 2.

4. PREPARATION OF SCATTER DIAGRAMS BY USE OF AN OSCILLOSCOPE

A major part of the work in using scatter diagrams to estimate correlation coefficients would ordinarily be the preparation of the diagrams. If the data under study were already tabulated then it is likely that it would be as convenient and more useful to compute the correlation coefficient directly by use of a desk calculator then to prepare the scatter diagram and analyze it. However if the variables under study are available in the form of varying voltages then it is possible to use an oscilloscope to plot the scatter diagram directly.[5] The oscilloscope is adjusted to give spot deflections in the horizontal and vertical directions proportional to the two variables x and y. A pulse generator is used to modulate the electron beam and produce dots of light on the oscilloscope screen at the desired sampling rate. If the screen is then photographed (time exposure) the desired scatter diagram is obtained. Figures 3 and 4 give examples of scatter diagrams made by this method for Gaussian and Rayleigh variables, respectively. The actual values of correlation coefficient and the estimates made by the method outlined above are given.

From a practical viewpoint it does not seem essential that a true scatter diagram be made. Examples of photographs made by exposure of film to a moving spot on an oscillograph (rather than a sequence of dots) are shown in Fig. 5. It appears that correlation can be estimated from these almost as well as from the corresponding scatter diagrams of Fig. 3.

5. ACCURACY OF THE PROPOSED METHOD

There are three main factors to be considered in assessing the accuracy of estimating correlation coefficients from scatter diagrams. Two of these are related to the statistics of the problem and the third to the mechanics of the method. Consider first the accuracy of determining the correlation coefficient from a set of

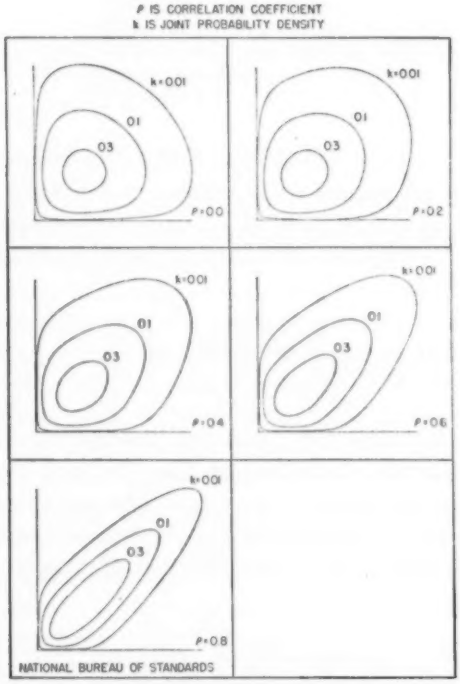


FIG. 2. Loci of constant joint probability density for normalized two-dimensional Rayleigh distributions.

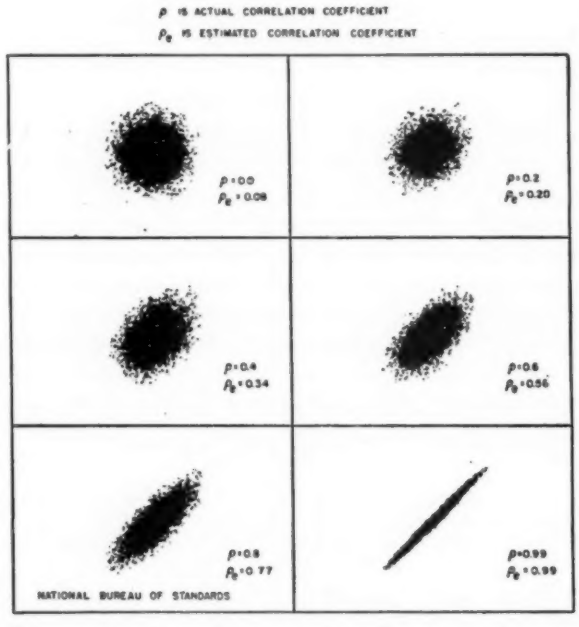


FIG. 3. Examples of scatter diagrams of two-dimensional Gaussian distributions.

[4] E. N. Bramley, J. Inst. Elec. Engrs. (London) III, 98, 19–25 (1951).

[5] J. C. R. Licklider and E. Dzendolet, Science 107, 121–124 (1948).

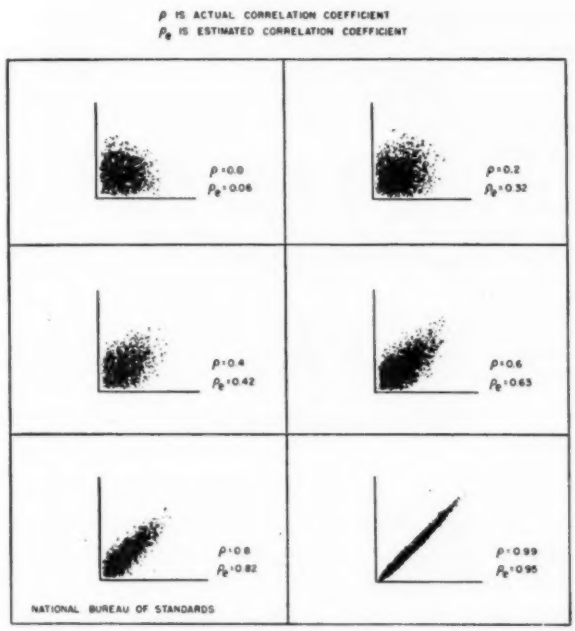


FIG. 4. Examples of scatter diagrams of two-dimensional Rayleigh distributions.

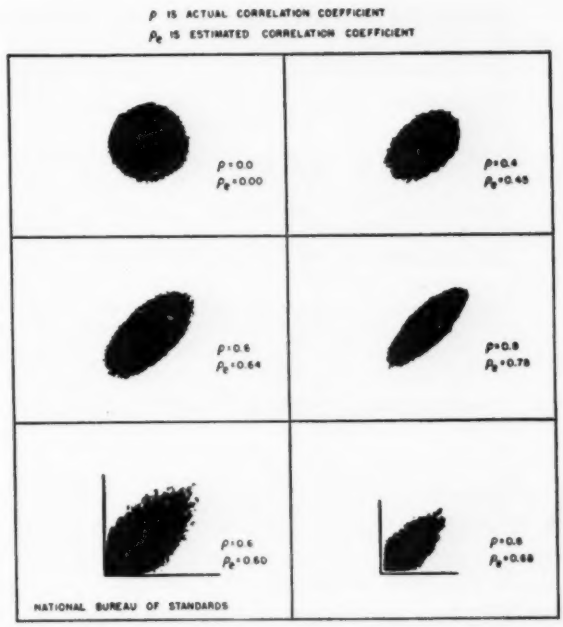


FIG. 5. Examples resembling scatter diagrams of two-dimensional Gaussian and Rayleigh distributions but made from continuous data.

tabulated data when it is assumed that a certain theoretical correlation exists. If n pairs are randomly selected from an infinite set of number-pairs, the correlation coefficient computed from these, and the process repeated many times, a random set of numbers ρ_i is obtained. Any single value of ρ_i is only an estimate of the true correlation coefficient. In the case where the two original variables are Gaussian distributed, the statistics of ρ_i are well known.[6,7] If it is desired that the error in estimating ρ by computation from a single set of n number-pairs be less than 0.05 then this will be accomplished with a probability ≥ 0.90 if $n \geq 1100$, with a probability ≥ 0.990 if $n \geq 2700$, and with a probability ≥ 0.9990 if $n \geq 4300$. It is believed that these same values may be used as an approximate guide in the case where Rayleigh distributed variables are considered.[8] However this case has not been investigated.

[6] H. Cramer, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, 1951), pp. 397–401.

[7] F. N. David, *Tables of the Correlation Coefficient* (*Biometrika* Office, University College, London, 1938).

[8] See reference 7, p. xvii.

The second statistical limitation concerns the distribution of the variables under study. Since the loci depend on this distribution then errors may arise if the form of it is not known. This source of error has not been investigated extensively. However as a rough check the loci for the Rayleigh case (shown in Fig. 2) were scaled as if the distribution were Gaussian. The ratio R was taken as the ratio of the maximum dimension in the direction $x=-y$ to that in the direction $x=y$. When scaled in this manner the largest error in determining ρ was -0.09 (estimate too small) and the median error was -0.05. In other cases the error from this source may be much greater.

The third source of error is the scaling of the scatter diagram. Since the loci of constant dot density are extimated by eye, subjective errors will arise. It is likely that this source of error will be the limiting one in practice and that errors in estimating ρ will be of the order of ± 0.1.

Fast Time Analysis of Intermittent Point-to-Plane Corona in Air. II. The Positive Pre-Onset Streamer Corona*

M. R. AMIN
Department of Physics, University of California, Berkeley, California
(Received August 3, 1953)

Analysis has been made of pre-onset streamers in room air at 760 mm with fast oscilloscope using both photomultiplier and electrical registration. The shape of the photon pulse observed by W. N. English is confirmed, and is interpreted. Current and luminosity roughly parallel each other in the streamer. Tip velocities ranged from 6×10^7 cm/sec to 2×10^7 cm/sec being highest near the point. The luminosity is constant over the path except at the very beginning and at the end of the path suggesting that ionization is also constant. The net number of positive ions left behind per cm length of streamer in consequence of the electron flow into the point is constant, and has been measured to be about 8×10^9 ions per cm in general agreement with earlier work. Lateral deflection of the ends of streamer tracks by space charge left behind from previous streamers was directly observed.

AS was indicated in a preceding paper,[1] certain very heavy pulses appear in the counting region after the burst pulse onset is reached and before the steady corona is achieved. These are the pre-onset streamer pulses. Streamer pulses were anticipated, in connection with positive point corona, by L. B. Loeb. Sought for in vain by G. W. Trichel in 1937, they were first observed and studied in the pre-onset region of the burst pulse corona by Kip.[2] Visually, they are characterized by being projected into the gap and giving the impression of a cone of luminosity. Kip studied their repeat rate and measured the number of ions in these pulses. Later on, the onset of streamers and various static studies about them were made by other workers.[3,4]

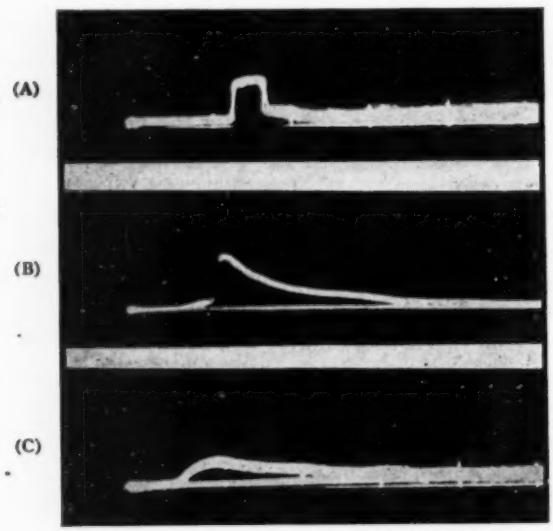


FIG. 1. Streamer pulses for 2.5-mm diameter point. 5-cm gap. $RC=5\times10^{-9}$ sec. (A) 2×10^{-8} sec test pulse. (B) Photon pulse looking at the whole gap. (C) Electrically induced streamer pulse.

In the case of the streamers, as in burst pulses instruments of inadequate time resolution prevented a thorough time study. In 1950 W. N. English,[5] using a fast micro-oscillograph and a photomultiplier, obtained one oscillogram of the streamer photon pulse which revealed the rapid time of formation and propagation of the streamers. The instrument was available for only one measurement, so that practically no data could be obtained beyond that for the photon pulse over the whole length of a streamer. It was therefore desirable to confirm and interpret the oscillogram of English and to extend his studies as far as the equipment available would permit. The equipment and techniques used were those indicated in an earlier paper.[1]

In the following all the studies were made in room air at atmospheric pressure. This was the most suitable medium since H. W. Bandel[4] has shown that the streamer formation is not favored in either dry air or at lower pressures. In all these studies, the Type 513D Tektronix Synchroscope was used. The shape of the pulses observed were the same for all the geometries studied and, therefore, only a few illustrative oscillograms for the 2.5-mm diameter point and 5-cm gap will be presented here.

Figure 1 traces (B) and (C) were obtained by direct application of both electrical and photon pulse through a delay line to the deflection plates of the cathode-ray tube of the synchroscope. When the photon pulse was under study, the synchroscope was externally triggered by the electrical pulse and vice versa. A rectangular test pulse of 2×10^{-8} sec was generated by discharging a charged-open circuit delay line through its characteristic impedance. The test pulse was then used for calibration. By feeding the test pulse through either the electrical system or the photomultiplier system to the deflection plates of the cathode-ray tube, the oscillogram of Fig. 1 trace (A) was obtained, showing the time resolution of the circuit.

In order to measure the velocity of the streamers and

* This work was supported for one year on funds from the U. S. Office of Naval Research and for a second year under a fellowship from the Research Corporation of America.
[1] M. R. Amin, J. Appl. Phys. (to be published).
[2] A. F. Kip, Phys. Rev. 55, 549 (1939).
[3] W. N. English, Phys. Rev. 74, 170 (1948).
[4] H. W. Bandel, Phys. Rev. 84, 92 (1951).
[5] W. N. English, Phys. Rev. 77, 850 (1950).

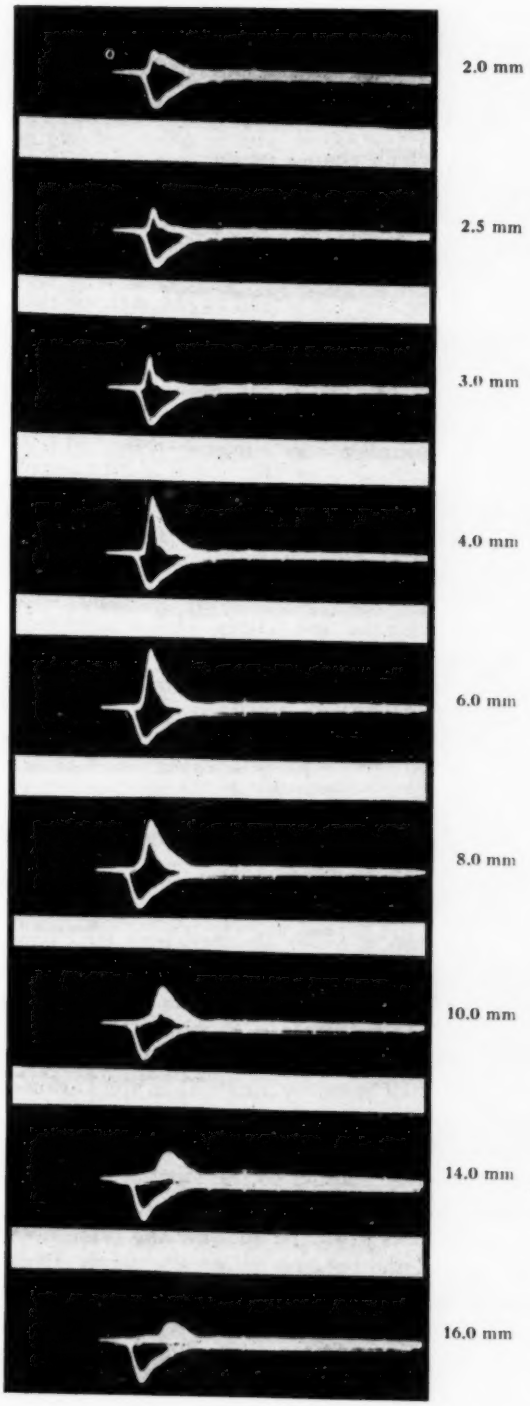


FIG. 2. Streamer pulses for 2.5-mm diameter point. 5-cm gap. $RC = 5 \times 10^{-9}$ sec. Sweep rate 0.1 μsec/scale division. Upper traces: Photon pulses from several 0.5 mm × 4 cm sections of the gap, at various distances from the point. Lower traces: Electrically induced pulses.

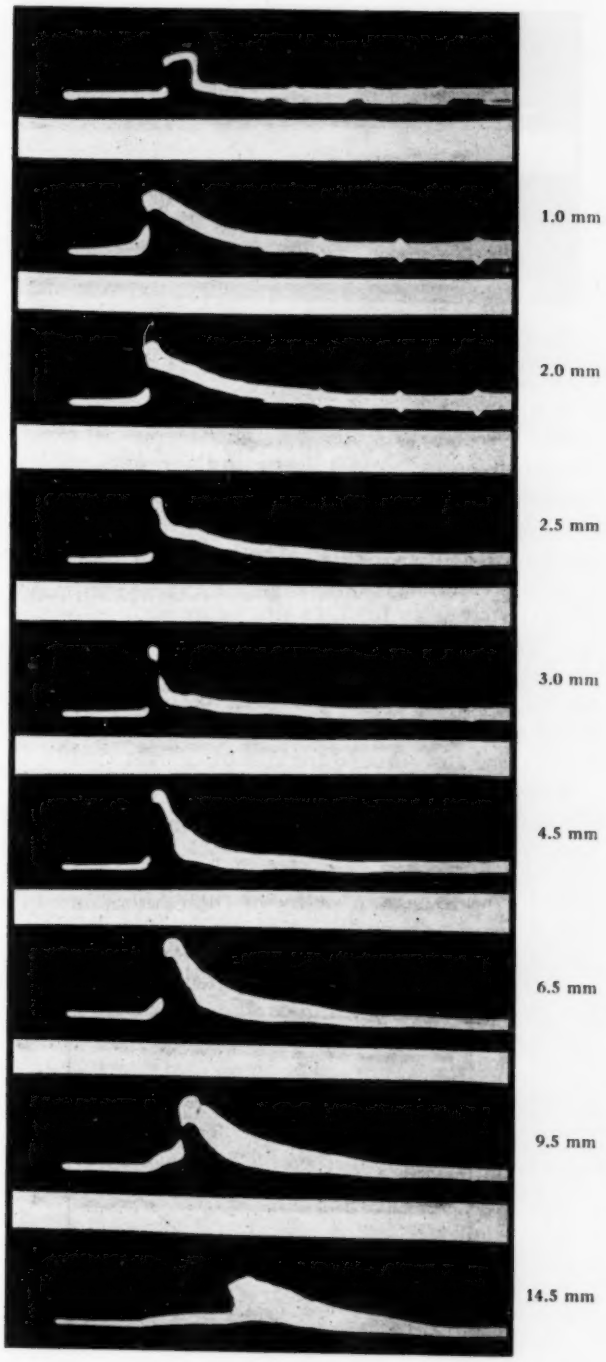


FIG. 3. Streamer pulses for 2.5-mm diameter point. 5-cm gap. $RC = 5 \times 10^{-9}$ sec. Uppermost picture: 2×10^{-8}-sec test pulse. Lower pictures: Photon pulses from several 0.5-mm by 4-cm sections of the gap at various distances from the point.

to be able to analyze the oscillograms of Fig. 1 traces (B) and (C), the following method was used. The gap between the point and the plane was focused on the photomultiplier. The slit system, in front and very close to the photomultiplier, was arranged so that a section of the gap 4 cm long in the horizontal direction and *only* 0.5 mm wide in the vertical direction was visible to the photomultiplier at one time. Keeping the dimensions of the slit constant and moving the photomultiplier housing smoothly in the vertical direction,

FIG. 4. Upper: Streamer photon pulse looking at the whole gap including the point. Lower: Streamer photon pulse looking at the whole gap excluding the point.

it was possible to study different 0.5-mm sections of the gap. Figures 2 and 3 illustrate the results.

Figure 2 was obtained by feeding the pulses through the amplifiers and the delay system of the synchroscope to the cathode-ray tube. The sweeping circuit was triggered externally in all cases by the electrical pulse (no external delay line). In this figure each oscillogram consists of two traces, the upper trace representing the photon pulse and the lower trace representing the electrical pulse. The sequence of oscillograms in this figure was taken as the slit was moved downward in the vertical direction to the distances indicated.

The amplifiers of the Tektronix, having a rise time of some 0.02 microsecond, distorted the pulses somewhat and, in order to study these pulses more accurately, they were directly applied (through an external delay line) to the deflection plates of the cathode-ray tube. This is illustrated in Fig. 3, showing a similar sequence of events as in Fig. 2, except that here it was not possible to get the electrical pulse on the same film. However, the sweeping circuit was triggered by the electrical pulse for all traces. The reproducibility of electrical pulses in Fig. 2 justifies time studies made from Fig. 3.

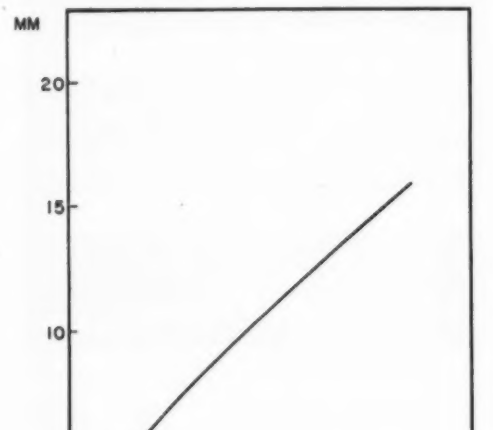


FIG. 5. Distance of streamer tip from the point *versus* time. 2.5-mm diameter point. 5-cm gap.

An item which should be emphasized at this point is that the traces in Figs. 1–3 are superpositions of many individual oscilloscope sweeps, thus showing the reproducibility of the individual events. Single streamers could be observed by the photomultiplier and thus were visually studied on the screen of the cathode-ray tube.

Owing to the lack of fast amplifiers some details, especially in the case of the electrical pulses, are missing. Caution must be used in making interpretations from the photon-pulse records, since in these observations the photomultiplier was drawing some 20-ma peak current which was well over its recommended linear response. Since the pulses were very short, the overload did not fatigue the photomultiplier and the observations are at least qualitatively correct. Another source of difficulty was the slow sweep rate of the synchroscope as compared to the fast pulses. This was remedied by taking the measurements from magnified oscillograms. The power supply and geometrical arrangement limited the study to the 4-cm and 5-cm gaps. The streamers produced in standard geometry were too short for this type of study.

1. Pulses Observed by the Photomultiplier

Regarding the photon pulse in Fig. 1, it is seen to consist of an initial slow rise, followed by a fast rise with a flat top, a lower short plateau which declines exponentially. This is the same sort of picture that English[5] observed, except that here the plateau is somewhat less prominent. This may be due to the different response of the RCA Type 1P21 (used in this work) and 5819 (used by English) to the various wave lengths.

The slow initial rise suggests some sort of small initial discharge leading to the streamer rise. In order to determine the location of this initial discharge the slit system was arranged so that the photomultiplier could see all the gap, *excluding* the point. The result is shown in Fig. 4. Here the upper oscillogram represents the whole gap *including* the point, whereas the lower oscillogram shows the streamer photon pulse, *excluding* the point. The initial slow rise is absent in the latter. Therefore, a reasonable conclusion is that the slow rise of luminosity takes place at the surface of the point. Since the streamers appear in the Geiger counter regime of burst pulses, it may be postulated that at the higher fields, the initial pulse of the burst pulse, while spreading over the point, perhaps aids in creating the narrow and dense region of ionization leading to the space charge that initiates the streamer. In fact, the magnitude and nature of the rise are quite similar to that for the initial burst pulse.

The streamer being triggered advances into the gap by virtue of its high-tip field and photoionization. Now, if the sharp spikes in Figs. 2 and 3 are identified as representing the passage of the tip of the streamer, the shift of the spike in time as the position of the slit is changed gives a measure of the velocity of propagation of the streamer tip. This identification is justified since it is expected that the luminosity is highest at, and just behind, the tip of the streamer. Thus, if the distance of the slit from the point is plotted against time, the slope of this curve indicates the velocity of propagation. Figure 5 is such a plot made from Fig. 3, giving a range of 2 to 6×10^7 cm/sec for the velocity of the streamer tip. Figure 2 gives almost the same plot. Nevertheless, Fig. 3 was chosen since it was more accurate. (The time-distance curves for various sized points and a 4-cm gap are given in Fig. 6.)

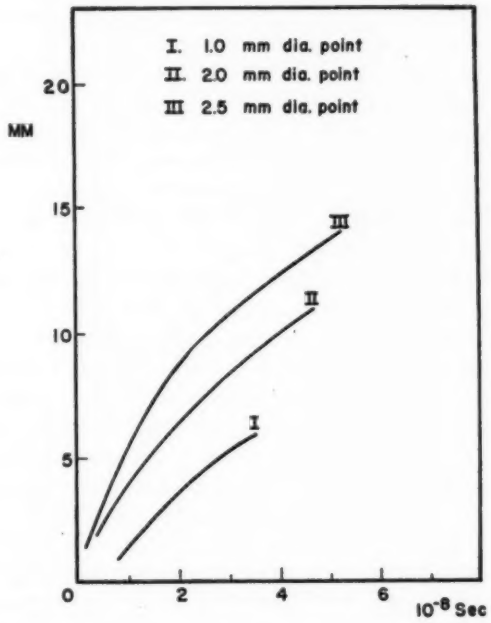


FIG. 6. Distance of streamer tip from the point *versus* time. Different sized points. 4-cm gap.

It is interesting to note that the velocity of these pre-onset streamers is high near the point and gradually decreases as the streamers advances. These streamers do not cross the gap and do not yield breakdown as they encounter too much space charge piled up out in the gap. The velocity of the breakdown streamers of Meek and Saxe[6] shows a behavior contrary to that of the pre-onset streamers, i.e., an initial slow advance and an increase in speed to a constant high value as the gap is crossed. This is to be expected since, in the case of breakdown streamers, the field is higher, and the space charge is removed as fast as it is formed. The extension of high field far into the gap by the breakdown streamer tip causes an increase in velocity.

[6] J. M. Meek and R. F. Saxe, Nature, **162**, 263 (1948).

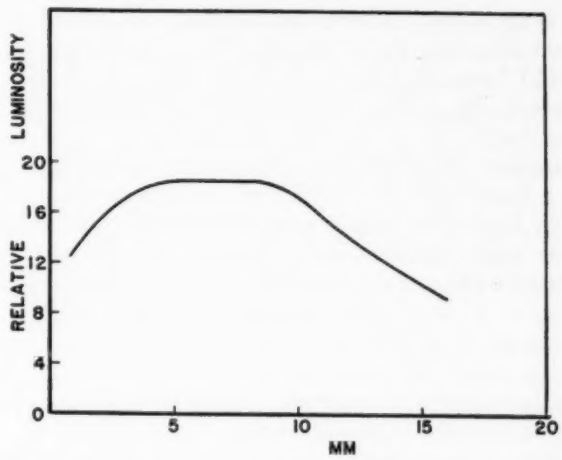


FIG. 7. Relative luminosity of streamer tip *versus* its distance from the point. 2.5-mm diameter point. 5-cm gap.

It is next of interest to plot the relative luminosity of the tip of the streamer (indicated by photon pulse height) as a function of position. This is shown in Figs. 7 and 8. The streamer tip in Fig. 7 reaches its maximum luminosity some 3 mm away from the point. In this 3 mm the streamer has a velocity of 6×10^7 cm/sec. This means that the peak of luminosity is reached in 5×10^{-9} sec, which is in agreement with the fast rise time of photon pulse in Fig. 1. Once the streamer has reached its maximum luminosity, it maintains this luminosity until it has advanced about 10 mm into the gap, Fig. 7. The velocity of streamer tip corresponding to this region is 3×10^7 cm/sec. Thus, when looking at the whole gap with the photomultiplier a plateau of luminosity lasting 2×10^{-8} sec is expected. This is indicated by the flat top of photon-pulse in Fig. 1. However, the flat top is shorter than 2×10^{-8} sec. This seeming discrepancy results from neglecting the luminosity in the streamer channel caused by after-effects which must be added to the tip peak intensity curve to give the composite, Fig. 1.

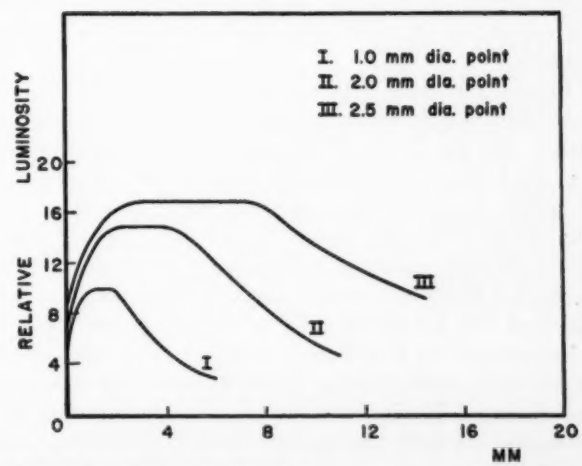


FIG. 8. Relative luminosity of streamer tip *versus* its distance from the point. Different sized points. 4-cm gap.

It is noted that following each peak at various distances from the point there is luminosity extending for some 5×10^{-8} sec after the peak in the form of a hump or plateau, Fig. 3. This plateau is more prominent initially and declines to zero as the streamer tip advances, say, 3 mm into the gap. If a rough estimate of this luminosity as a function of time is plotted and superposed on the peak luminosity-time curve† the resulting luminosity-time curve for the whole streamer will very closely resemble the one shown in Fig. 1, with a narrow peak. Since it is very difficult quantitatively to estimate the luminosity-time curve for the whole streamer path owing to the luminosity following the streamer tip, the synthetic curve of luminosity *versus* time composed of the tip luminosity and after-effect is not presented.

Now, if one assumes that excitation and ionization parallel each other at the tip of the streamer, Figs. 7 and 8 give the relative density of ionization per cm along the streamer path. Accordingly, one may conclude that the total ionization and excitation along the path of the streamer is nearly constant, except for the short distance near the point and the somewhat longer distance at the end of the streamer path. The assumption of parallelism of excitation and ionization is justified as will be noted by quite a number of observations comparing electrical and photon pulses in all these investigations. It appears also that the excitation and ionization do vary among streamers from different points. However, measurements of total number of ions in the streamer pulse indicate that the net number of positive ions left behind in the streamer channel, after the electron current to the point has ceased, are relatively constant when divided by the streamer lengths. That is, the net number of positive ions left behind per cm for the streamers II and III of Fig. 8 are nearly the same, while those for I are slightly less. Similar conclusions had been drawn by Kip[2] and English[3] from electrical measurements over a large range of streamer lengths and confirmation comes from the work of Meek and Saxe.[6] It is necessary at this point merely to indicate the relationship and to leave the quantities for discussion under the electrical pulse analysis.

Figure 3 reveals that as the streamer advances, a slow rise appears which is very prominent in the lower traces. On the basis of streamer mechanism and the fact that this corresponds in time to the appearance of spikes of luminosity near the point, this slow rise was identified as luminous section ahead of the streamer tip, probably produced by photoexcitation and some ionization of the gas by the intense light of the streamer tip, and the advanced field of the streamer. This is substantiated by Kip's[2] early study of streamers where intense photoionization in the gas ahead of the streamer was observed. For it is not unlikely that as the streamer advances, its field lowers but extends further out by space charge expansion. A similar observation and interpretation on the basis of such excitation was made by Meek and Saxe[6] for breakdown streamers.

Figure 3 shows that near the point the relatively sharp spike is followed by a plateau. This plateau gradually decreases as the streamer advances into the gap. The plateau indicates a luminous phenomenon which takes place after the streamer tip has crossed the observation slit. This plateau is responsible for the rather peculiar shape of the photon pulse when the whole gap is under observation, Fig. 1.

L. B. Loeb, on the basis of his recent studies of these streamers and those of lightning strokes, has indicated that the positive space charge per cm streamer length left behind by the streamer tip through a flow of electrons up the streamer, is so heavy that a radial field causes ionization by converging electrons. This process will take place for some 5×10^{-8} sec after the peak passes. Thus, it could very well cause after-luminosity reflected in the plateau of English[3] and the plateau observed here. Recombination in 5×10^{-8} sec would be too small to account for all of this light, though it may contribute some. With the fast scale of events here, it must be noted that the average radiation time of excited states is of the order of 10^{-8} second and, thus, must be reckoned with in these interpretations. The radial expansion might be out to perhaps 0.5 mm. Because of the limited resolution of the slit system, this expansion, in any case, could not be detected.

Previous observers could only estimate streamer length by visual observations (Kip[2] and English[3]). With the slit system of the photomultiplier, the length of the streamer could be directly measured. The slit system was arranged so that a 0.5-mm × 0.5-mm area of the gap was focused on the photomultiplier. The slit was then moved in a vertical direction parallel to the axis. Special care was taken so that only the streamers in the axial directions were studied, thus avoiding the streamers which had a horizontal component. The distance at which no sharp spike appeared was designated as the length of the streamer. This method gave smaller values for the streamer length than those of Kip and English. Table I gives the streamer length for different sized points. With the above slit arrangement it was also possible to detect the distance of deflection of some of the streamers from the axis. This was done by moving the photomultiplier housing in the horizontal direction. It appeared that near the point all the streamers fell in a channel less than 1.0 mm wide, whereas, near the end of their path they were detected as far as 7 mm from the axis. The deflection of individual streamers is caused by the presence of space charge in the relatively low field region created by the preceding streamers and burst pulses. Individual streamer pulses can thus be seen deflecting this way and that. It is this behavior which gives the impression of a luminous pencil ending in a shaving-brush-like

† This can be directly obtained from Figs. 5 and 7 by elimination of distance.

configuration when streamers are observed visually. The positive ions in the high field region are removed very fast, and a second streamer cannot start until the space charge is removed to a distance far enough to restore the necessary field for streamer propagation. Thus, the streamers follow the axial direction in the beginning of their path and are deflected whenever they encounter the space charge. It is this deflection which gives the smearing of the photon pulses of Fig. 3 at a distance of, say, 10 mm from the point. It is not impossible that the reappearance of the hump beyond 4.5 mm in Fig. 3 is in part caused by recombination of ions as the current declines and the streamer ceases. While the current flows in the channel recombination cannot occur effectively. When it ceases it can. However, probably most of the smearing of luminosity is caused by the lateral deflection of some of the streamers as viewed by the long slit.

2. The Electrical Pulse

The electrical pulse, because of its short duration, could not be amplified satisfactorily with the existing amplifiers. When applied to the deflection plates of the cathode-ray tube through a delay line, it appeared as trace C of Fig. 1. The induced current has a peak value of *0.06 amps*. To measure the number of ions in a streamer pulse, the current pulse was integrated over the time of duration, neglecting the small contribution due to the motion of positive ions in low field region. This approximation is justified by the Green's reciprocity theorem. To check this calculation, the screen method of Kip[2] was used. The results are in fairly good agreement. Whenever a streamer is followed by a burst pulse the screen method may have an error of perhaps 30 percent, since the number of ions of burst pulse are also included in the measurement. Table I gives the number of ions per streamer for different sized points.

Assuming that the net positive space charge is constant along the path of the streamer, Table I gives the values of the net number of positive ions/cm of the streamer. The average value is near 8×10^9 per cm length of the streamer. These values are slightly larger than those of Kip. This results from the fact that in his investigation the length of the streamer was estimated from visual observation, thus giving a larger value than actually measured here. However, the total number of ions in a streamer for the same geometrical arrangement is in good agreement with the results of Kip and English.

SUMMARY

Despite the limitations indicated, the investigations confirmed the shape of the photon pulse observed by English. They further revealed that the pre-onset streamer is initiated by a pulse akin to the initial pulse in a burst pulse. They reveal that the current and luminosity roughly parallel each other‡ and that the streamer tip is characterized by a sharp spike of luminosity which advances across the gap. Its initial velocity is high (6×10^7 cm/sec) but declines to 2×10^7 cm/sec as the tip advances to the end of its travel. The streamers travel along a fairly straight path following the axis for most of their path. Eventually they encounter the positive space charge of a preceding streamer and are deflected, to one side or the other, producing the familiar brush visually observed. The oscillograms give evidence of excitation and ionization for a short distance behind the tip, caused by radial expansion of the net positive space charge created, as suggested by Loeb. The streamer creates a net space charge of positive ions of 8×10^9 ions per cm length, which is fairly constant except at the beginning and end of the path confirming previous inferences. This entails an electron current up the channel of about 0.025 ampere during the advance.

As L. B. Loeb[7] has recently shown, the removal of electrons and the leaving behind of a net positive space charge n of 8×10^9 ions per cm of streamer together with tip velocities v_t from 2×10^7 to 6×10^7 cm/sec requires that the electron current be carried by n_x electrons traveling with a drift velocity v_x, compatible with potential gradients down the channel, such that $v_x=n/n_xv_t$ with n_x considerably greater than n. Thus the number of ions created per cm length of streamer n_x+n is indeed high. The oscillograms further indicate that as the streamer begins to slow down its tip decreases in height and broadens. Much photoionization and excitation can then be seen in advance of the tip, and there is evidence that some luminosity is created by recombination behind the peak as it slows down. Owing to lack of resolution the photomultiplier-detected diameter of the streamer channel can only be set as less than 0.5 mm, and it can well be as little as 0.1 mm.

TABLE I

Point diam. in cm	0.1	0.2	0.25
Streamer length in cm	0.60	1.10	1.45
Number of ions per streamer	3.5×10^9	9.0×10^9	1.2×10^{10}
Number of ions per cm length of the streamer	5.8×10^9	8.1×10^9	8.3×10^9

The writer is indebted to Professor L. B. Loeb for suggesting the problem, and for his continued interest and valuable assistance in interpretation of some of the more complex phenomena observed. He also wishes to thank Mr. H. W. Bandel for his kind assistance in the beginning of the work.

‡ This fact was more strikingly shown in study of the Trichel pulses, where amplification of electrical and photon pulses was comparable. In the present study the electrical pulse could not be amplified under conditions making electrical and photon pulses as accurately comparable.

[7] L. B. Loeb, Phys. Rev. (to be published).

Diffusion of Gold into Copper*

A. B. MARTIN, R. D. JOHNSON,† AND FRANK ASARO‡
North American Aviation, Inc., Atomic Energy Research Department, Downey, California
(Received August 5, 1953)

Radioactive tracer techniques have been utilized to measure the rate of diffusion of gold into polycrystalline copper over the temperature range from 1000° to 375°C. These experiments have yielded the values $Q=44\,900\pm1300$ cal/mole and $D_0=0.10\pm0.06$ cm²/sec for the volume diffusion process in the temperature range from 1000°C to 750°C. Diffusion measurements at 550°C and 375°C indicated that the measured rates at these temperatures had been enhanced by grain-boundary diffusion. Considerable attention was directed toward the variation of the diffusion coefficient with (1) the purity of the copper, (2) the grain size of the copper, (3) the thickness of the plating, and (4) the time of anneal. Variations of these parameters showed no significant effect on the diffusion coefficient, with the exception of the time of anneal. The measured diffusion rate at 1000°C was observed to decrease with increasing time of anneal.

INTRODUCTION

STUDIES of the diffusion of gold into copper have been reported by Matano,[1] who used an x-ray method to follow the progress of diffusion from a copper-rich gold alloy into a thin layer of pure copper. Although the data did not accurately fit the Arrhenius equation, an average activation energy of 22 450 cal/mole was derived for the entire temperature range investigated. For the temperature range from 650°C to 970°C, the activation energy was stated to be 27 400 cal/mole. When these results are compared with activation energies for self-diffusion in copper (45 to 61.4 kcal/mole) it is seen that they are in agreement with the conclusion that has been widely accepted in the past that the heat of activation involved in the chemical diffusion of a dilute solute is considerably less than that for self-diffusion of the pure solvent.

More recently, however, Nowick[2] has reinterpreted the data from a number of intermetallic diffusion experiments and has suggested that activation energies for chemical diffusion should be very nearly the same as that for self-diffusion. The low values of the activation energy, Q, and the frequency factor, D_0, which have been measured in the past for chemical diffusion are attributed to the fact that solute atoms have a greater preference for short-circuiting paths (such as grain boundaries) or to an alternative possibility that diffusion under a chemical gradient produces strains which in turn produce additional short circuiting paths. The extent to which grain-boundary diffusion lowers the measured activation energy is greater the lower the temperature of the measurement.

Nowick's suggestion has received support from the recent work of Slifkin and co-workers.[3] Using radioactive tracer methods, the activation energy for the diffusion of antimony into silver was found to be 39.4 kcal/mole, which differs greatly from the older value of 21.7 kcal/mole reported by Seith and Peretti.[4] The heat of activation for self-diffusion in silver is 45.9 kcal/mole.[5] Tomizuka, Slifkin, and Lazarus[6] recently measured the activation energy for diffusion of cadmium into silver to be 41 kcal/mole. The older value was 24.8 kcal/mole.[4]

The experiments to be reported here were designed to measure the rate of diffusion of gold into copper under well-controlled experimental conditions and to observe the effect on the measured diffusion rate of variations in several parameters, namely the purity of the solvent, grain size of the solvent, thickness of initial solute plating, and time of anneal. In order to extend the measurements to relatively low temperatures and long times of anneal a technique somewhat different from the conventional radioactive tracer method for intermetallic diffusion experiments was employed.

EXPERIMENTAL METHOD

In the conventional tracer experiments a thin layer of solute, containing the radioactive atoms, is electroplated or otherwise deposited on the face of a cylindrical bar of the solvent material. After the diffusion anneal, the penetration of the solute is determined by sectioning the sample and measuring the radioactivity in each cut. However, this technique is suitable only when the half-life of the radioactive solute atoms is long enough for a measurable amount of activity to remain after the diffusion anneal. This limitation quite often prevents carrying out diffusion anneals that are long enough to obtain measurable penetrations at the lower temperatures. This limitation exists in the case of diffusion of radioactive Au^{198} (2.7-day half-life) into copper.

* This paper is based on studies conducted for the U. S. Atomic Energy Commission under contract AT-11-1-GEN-8.
† Now at Battelle Memorial Institute.
‡ Now a U. S. Atomic Energy Commission fellow at University of California, at Berkeley.
[1] C. Matano, Japan. J. Phys. 9, 41 (1933).
[2] A. S. Nowick, J. Appl. Phys. 221, 182 (1951).
[3] Slifkin, Lazarus, and Tomizuka, J. Appl. Phys. 23, 1405 (1952).
[4] W. Seith and E. Peretti, Z. Electrochem. 42, 570 (1936). See also C. J. Smithells, *Metals Reference Book* (Interscience Publishers, Inc., New York, 1949).
[5] Slifkin, Lazarus, and Tomizuka, J. Appl. Phys. 23, 1032 (1952).
[6] Tomizuka, Slifkin, and Lazarus, Am. Phys. Soc. Bulletin 28, No. 2, Paper Z-10 (March 26, 1953).

To overcome this problem the diffusion specimens for long-time anneals were prepared by electroplating a thin layer of stable gold on the flat end of a pure copper rod and, after a diffusion anneal of approximately two months, the specimens were activated by a 16-day exposure to neutron irradiation in a cooled test hole of a reactor. This exposure was sufficient to activate the gold to within one percent of saturation for the flux present. After irradiation, the specimens were allowed to cool for a sufficient length of time for the activated copper isotopes, with a maximum half-life of 12.8 hours, to decay to essentially zero activity, and only the activity from the 65-hour half-life gold isotope Au^{198} remained. The specimens could then be processed in the usual manner.

This method requires the use of very pure materials for both solute and solvent to minimize the effect of extraneous activities which may result from small amounts of impurities which have high neutron capture cross sections. The technique is limited, of course, to cases in which the radioactivity resulting from neutron capture in the solvent material has a significantly shorter half-life than the activity produced in the solute.

For all of the samples except those subjected to the longest anneals (59 days), the conventional tracer method was employed.

EXPERIMENTAL PROCEDURES

A. Materials

Radioactive gold was obtained from the Isotopes Division of the U. S. Atomic Energy Commission. The rate of decay of the activity was measured to establish the absence of radioactive impurities. A 65-hour half-life was observed, which is in good agreement with the half-life for Au^{198} reported by other workers.[7] The stable gold used was 99.999 percent, obtained from Johnson, Matthey, and Company. Two grades of polycrystalline copper were used: 99.999 percent, obtained from Johnson, Matthey, and Company in the form of rods 7 mm in diameter, and American Brass 99.95 percent, in one-quarter-inch diameter rods. In all cases in which the specimens were activated by neutron irradiation, both the solute and solvent materials were 99.999 percent pure. The two grain sizes employed were the "as received" grain size, referred to as "small," and a stabilized grain size of approximately 1-mm mean linear dimension obtained by a 30-hour anneal at 1000°C, hereafter referred to as "large."

B. Preparation of the Diffusion Couples

Quarter-inch long pieces of the copper rods were used for the solvent. After the grain stabilization, when employed, the rods were mounted in Lucite disks and polished. The stable (carrier) gold and the radioactive Au^{198} were electroplated from potassium cyanide solution. The entire chemical system for transferring the radioactive solution from storage to the Lucite plating cell was enclosed in temporary lead boxes in a chemical hood and provided with remote controls.

Each couple of Au^{198} and copper was plated until the radiation intensity at one-half centimeter from the gold surface was approximately 100 mr/hr. Then a few mils were trimmed off the cylindrical periphery of the sample to remove any gold that might have been plated on the sides of the cylinder.

C. Processing the Diffusion Specimens

For the diffusion anneals, the samples were sealed off in evacuated Vycor tubes. The heat treatments were carried out in a small muffle furnace in which the temperature of the specimens was held constant within $\pm 2C°$ by an automatic recorder-controller. The Vycor tubes were air quenched at the end of the diffusion anneal. The low temperature specimens were furnace annealed for 59 days at temperature of 375°C and 550°C. They were then subjected to neutron irradiation in a reactor in a cooled test hole in which the temperature was maintained near room temperature.

Prior to sectioning, the cylindrical surface of each specimen was turned off to a depth of a few mils as a precaution against the possibility of surface diffusion or vapor phase transport of radioactive material during the anneal. The sample was then sectioned by taking successive cuts of approximately two mils in thickness (less in the cases of low penetration). The chips from each cut were collected, weighed, dissolved in hot aqua regia, diluted to a standard volume, and measured for radioactivity by means of a dipping Geiger counter. Aliquots were taken in order to obtain approximately the same counting rate for samples taken from a wide range of specific activities. Solution counting is particularly convenient in the case of samples diffused at the lower temperatures where reasonable annealing times still yield very small penetrations.

The importance of using pure materials in neutron activated diffusion couples is illustrated by the fact that, even though the copper employed as the solvent in these experiments was reported to be 99.999 percent pure, a small amount of long-lived activity was observed in all of the specimens which were activated. This activity was present throughout the entire specimen. Hence, the activity was associated with either the copper solvent or an impurity in the solvent. Although a positive identification of this activity was unsuccessful, a decay measurement indicated a half-life of approximately 190 days. Therefore, the activity measurements were corrected for this background activity by subtracting the activity determined from

[7] G. T. Seaborg and I. Pearlman, Revs. Modern Phys. **20**, 585 (1948).

a section taken off the opposite end of the diffusion specimen from the gold-plated end. The decay curve for the measured activity minus the background activity yielded approximately the same slope for eight sections of one of the specimens. These slopes corresponded to half-lives of 60 to 70 hours as compared with the accepted value of 64.8 hours for the decay of Au^{198}.

From chemical analysis of the gold in the first section removed from the diffusion couple and the relative activities in each section the amount of gold in each section was computed. Then the calculated values of x, the distance from the plated end to the center of mass of each section, computed from the total weight of the sections and the density of copper, were corrected for the amount of gold in each section, assuming a linear variation of density with composition. Also the values of x were corrected to the original interface since the total thickness of the plating was known from the integrated gold activity. The values of x were also corrected for the temperature of the anneal by using the linear coefficient of expansion for copper.

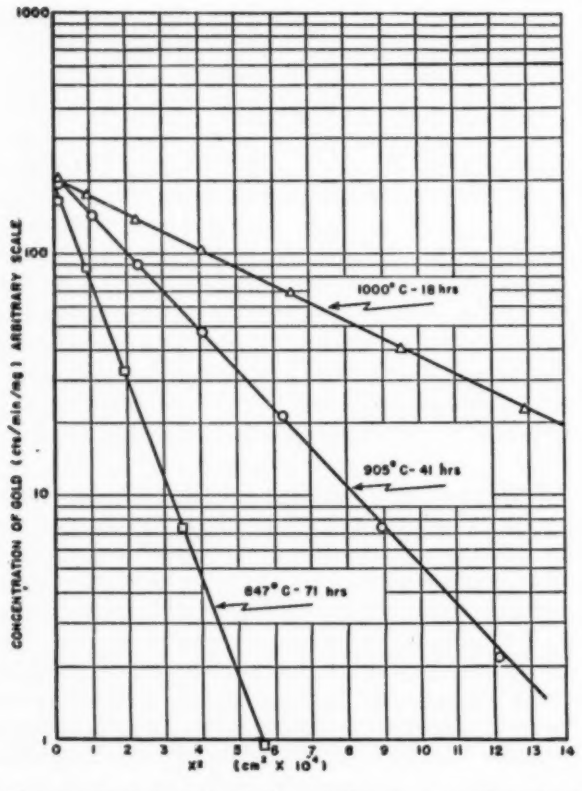

1000
100
10
1
CONCENTRATION OF GOLD (cts/min/mg) ARBITRARY SCALE
1000° C - 18 hrs
905° C - 41 hrs
847° C - 71 hrs
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
x² (cm² X 10⁴)

FIG. 1. Typical plots of the relative concentration of gold *versus* the square of the distance of penetration from the original interface. The linearity of these data lends support to the assumption that the diffusion coefficient is concentration-independent.

EXPERIMENTAL RESULTS

The method of analyzing the data has been described completely by W. A. Johnson.[8] Stated briefly, it consists of solving Fick's equation for diffusion in one direction with the boundary conditions implied by a thin plating of solute on the flat end of a cylindrical rod and with the assumption that the diffusion coefficient, D, is concentration-independent. The solution for the concentration, c, of the solute at a distance, x, from the original interface after a time, t, can be written as

$$\ln c = -x^2/(4Dt) + \ln[c_0/2(\pi Dt)^{\frac{1}{2}}], \qquad (1)$$

where c_0 is a constant proportional to the amount of solute plated on the solvent. If the logarithm of the concentration is plotted against the square of the distance from the original interface, the resulting plot is a straight line of slope $-1/(4Dt)$. From the measured slope and the time of anneal, the diffusion coefficient at the temperature of the anneal can be readily computed. A detailed discussion of the errors involved in the measurement of D with these radioactive tracer methods and analysis is discussed elsewhere.[9]

Typical plots of the logarithm of the concentration *versus* the square of the penetration are shown in Fig. 1. The linearity of these plots supports the assumption that was made concerning the concentration-independence of the diffusion coefficient. The variation of the diffusion coefficient with temperature is shown in Fig. 2, in which the logarithm of D is plotted against the reciprocal of the absolute temperature. Because of the number of variables involved, the complete data for the various diffusion anneals has been tabulated in Table I. As a result of the large amount of grain-boundary diffusion which occurred in the anneals at 375°C and 550°C, the log c *versus* x^2 plots were not straight lines. Therefore D values were computed from the straight lines which were drawn as the initial slopes of these curves. These slopes usually passed through the first three or four data points on the curve. Admittedly, these measured diffusion coefficients have been enhanced by grain-boundary diffusion although no differences were detected in the measured diffusion coefficients in the large- and small-grained solvents.

The independence of D of initial grain size for these cases, where grain-boundary diffusion is important, indicates that the "mean" grain size during diffusion in the "as received" samples was almost as large as the "stabilized" grain size.

The diffusion coefficients in the temperature range from 750°C to 1000°C were fitted to the Arrhenius law by the least squares method in order to determine D_0, the frequency factor, and Q, the activation heat of

[8] W. A. Johnson, Trans. Am. Inst. Mining Met. Engrs. **143**, 107 (1944).

[9] R. D. Johnson and A. B. Martin, J. Appl. Phys. **23**, 1245 (1952).

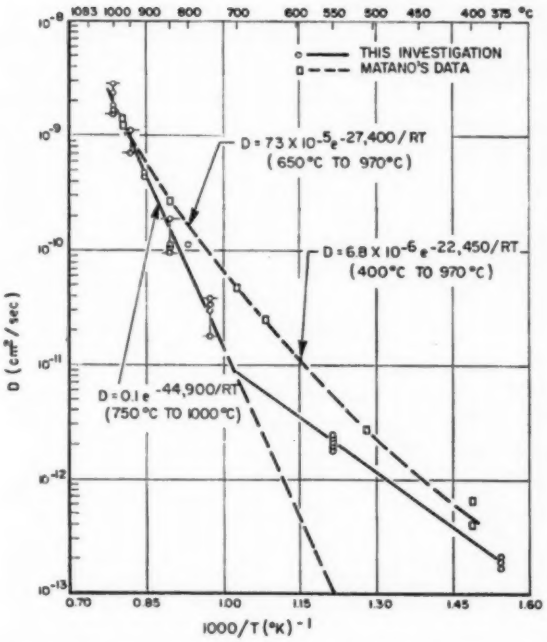


FIG. 2. Log D *vs* $1/T$ for diffusion of gold into copper. The diffusion coefficients measured at the lower temperatures, 550°C and 375°C, have been enhanced by relatively large amounts of grain-boundary diffusion. The intersection of the two curves drawn through the present data indicates that 600°C–700°C is the temperature range in which the amount of matter diffusing along grain boundaries becomes comparable to the amount transported by volume diffusion.

diffusion. The values obtained were:

$D_0 = 0.10 \pm 0.06$ cm²/sec and $Q = 44\,900 \pm 1300$ cal/mole.

The errors indicated are the probable errors calculated by the least squares method.

Although the least squares method of analyzing the data yields a relatively small probable error in the values of D_0 and Q, Fig. 2 shows a scatter in the D values measured at a given temperature that is too large to be accounted for by the known uncertainties in the various experimental factors involved in these measurements. The variations in the D values at a given temperature correlate most consistently with the time of anneal, the measured value of D being smaller the longer the annealing time (see in particular the data at 1000°C and 750°C in Fig. 2 and Table I). This behavior would normally imply that D is concentration-dependent. However, no correlation could be obtained between the known concentrations of gold and the observed variations in D. Furthermore, attempts to correlate the scatter in D at a given temperature with either the grain size or the copper purity were also unsuccessful. Therefore, experiments were designed to vary the thickness of the plating by a factor of 50, which in turn varied the concentration, and also to vary the time of anneal at 1000°C.

Specimens were prepared from two batches of radioactive gold which differed in specific activity by a factor 50, but all samples were plated with the same amount of radioactive gold. For each of two annealing times, 8 hours and 93 hours, half of the six specimens had platings which were 50 times thicker than the platings on the remaining specimens. The thicker platings resulted in correspondingly higher gold concentrations. The measured D values after 1000°C anneals are plotted against the time of anneal in Fig. 3 along with the data for 6-, 18-, and 30-hour anneals from Table I. It can be seen that there is no correlation of the observed variation in D with the thickness of plating or with gold concentration. However, the diffusion coefficient apparently decreases with increasing time of anneal up to 30 hours, and is approximately constant from 30 hours to 93 hours. Unfortunately, this effect is not inherent in diffusion theory, and a reasonable explanation for it is still being sought.

DISCUSSION

The results of this investigation are compared with the data of Matano in Fig. 2. Matano studied the diffusion of gold from a copper-rich gold alloy (roughly 90 atomic percent copper) into a thin electrodeposited layer of pure copper, the extent of diffusion being determined from lattice constant measurements. He attributed the nonlinear plot of log D *versus* $1/T$ to recrystallization, since the grain size of his diffusion specimens had not been stabilized prior to the diffusion anneal. However, the x-ray method measures the total amount of diffusion taking place regardless of whether it occurs by volume or grain-boundary diffusion.

TABLE I. Experimental data for gold-copper diffusion.

Grain size	Copper purity (%)	Duration of anneal (hours)	Temperature (°C)	Diffusion coefficient (cm²/sec)
Small	99.999	41	905	4.43×10^{-10}
Small	99.999	41	905	4.50×10^{-10}
Small	99.999	30	1000	1.50×10^{-9}
Small	99.999	30	1000	1.56×10^{-9}
Small	99.999	71	850	1.12×10^{-10}
Small	99.999	71	850	0.94×10^{-10}
Small	99.999	94	750	3.79×10^{-11}
Small	99.999	94	750	3.14×10^{-11}
Large	99.999	18	1000	2.31×10^{-9}
Large	99.999	6	1000	2.84×10^{-9}
Large	99.999	20	950	0.69×10^{-9}
Large	99.999	59.5	850	1.92×10^{-10}
Large	99.999	164	750	3.08×10^{-11}
Small	99.95	18	999	1.79×10^{-9}
Small	99.95	20.3	947	1.12×10^{-9}
Small	99.95	41	905	4.69×10^{-10}
Small	99.95	71	847	1.05×10^{-10}
Small	99.95	120	800	1.15×10^{-10}
Small	99.95	235	750	1.73×10^{-11}
Small	99.999	1416	375	2.13×10^{-13}
Small	99.999	1416	375	1.71×10^{-13}
Large	99.999	1416	375	1.97×10^{-13}
Small	99.999	1416	550	2.48×10^{-12}
Small	99.999	1416	550	2.10×10^{-12}
Small	99.999	1416	550	1.80×10^{-12}
Small	99.999	1416	550	1.92×10^{-12}
Large	99.999	1416	550	2.30×10^{-12}
Large	99.999	1416	550	2.15×10^{-12}

TABLE II. Experimental data for copper self-diffusion.

Investigators	Year	Q(kcal/mole) [b]	D_0 [b] (cm²/sec)	Temperature range (°C)	Method
[a] Rhines and Mehl Reference 11	1938	40	0.04	700–900	Chemical analysis and sectioning techniques.
Rollin Reference 12	1939	61.4	47	830–1030	Radioactive tracer with sectioning techniques.
Steigman, Shockley, and Nix, reference 13	1939	57.2	11	750–950	Radioactive tracer with face-counting techniques.
Raynor, Thomassen, and Rouse, reference 14	1942	46.8	0.3	650–850	Radioactive tracer with face-counting techniques.
Maier and Nelson Reference 15	1942	45.1	0.1 (polycrystals)	802–993	Radioactive tracer with sectioning techniques.
			0.6 (single crystals)	848–977	
Kuczynski Reference 16	1949	56	70	400–1000	Sintering of metallic particles.
Dedrick and Gerds Reference 17	1949	55	0.12	700–1025	Sintering of metallic particles.
[a] This investigation		44.9	0.1	750–1000	Radioactive tracer with sectioning techniques.

[a] These data pertain to rates of diffusion of various solutes in copper at low concentrations.
[b] The values of Q and D_0 have been determined from the slope and intercept, respectively, of log D *vs* $1/T$ plots rather than from any empirical formulas relating D, D_0, and Q.

Matano obtained an average activation energy of 22 450 cal/mole over the entire temperature range or 27 400 cal/mole over the temperature range from 650°C to 970°C. However, at the highest temperatures where grain-boundary diffusion would be relatively unimportant, the data are in agreement with the present work.

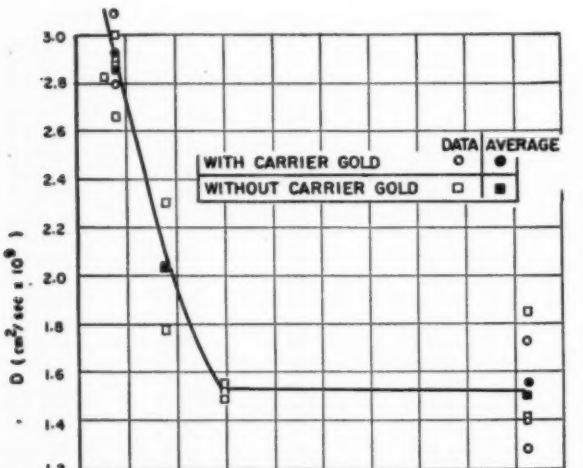


FIG. 3. Variation of the diffusion coefficient with time of anneal at 1000°C. The original plating of gold on a portion of the samples (○) was 50 times thicker than on a second portion (□), resulting in higher solute concentrations in the first set (○). There is no clear-cut correlation of the variation in the diffusion coefficient with either thickness of plating or solute concentration.

In the work reported here the measured D values at 375°C and 550°C are considerably higher than the extrapolated values from the high-temperature range. Apparently the D values at the lower temperatures have been greatly enhanced by the grain-boundary diffusion process. They may also be high because of the apparent dependence of D on the time of anneal. The intersection of the two straight line plots in Fig. 2 at a temperature of about 700°C should be a good indication of the temperature at which the grain-boundary diffusion process, which has a lower activation energy than volume diffusion, has an important effect on the measured volume diffusion coefficients in polycrystalline solvents. This crossover point is in the same temperature range as the 660°C temperature for the knee of the log D *versus* $1/T$ curve which was determined for self-diffusion in polycrystalline silver by Hoffman and Turnbull.[10]

It was also of interest to compare these results with the data of Rhines and Mehl,[11] who measured the diffusion coefficients of various metals (Zn, Al, Be, Cd, Si, and Sn) into copper and extrapolated their results to zero concentration of solute. The resulting ln$D$ *versus* $1/T$ plots were all similar and the authors suggested that, for the limiting case of very small concentrations, the data approached the self-diffusion coefficients for copper. If one excludes the data for Be, which has the smallest atomic radius of the solutes studied, and that for Cd, for which the data were taken in a lower tem-

[10] R. E. Hoffman and D. Turnbull, J. Appl. Phys. **22**, 634 (1951).
[11] F. N. Rhines and R. F. Mehl, Trans. Am. Inst. Mining Met. Engrs. **128**, 185 (1938).

perature range, the activation heats of diffusion are nearly the same and average about 40 000 cal/mole. This value is somewhat less than the 44 900 cal/mole determined in the present work.

For the purpose of comparison, the available experimental data for self-diffusion in copper[12–17] have been tabulated in Table II together with the average values of the data of Rhines and Mehl and the results of this investigation, which pertain to the rates of diffusion of various solutes in copper at low solute concentrations. It is interesting to note that the results of Maier and Nelson[15] for self-diffusion in polycrystalline copper are very nearly the same as the coefficients determined in the present work for the diffusion of gold into copper. It is perhaps significant that the experimental methods used in these two investigations were almost identical.

The present results are consistent with the features of intermetallic diffusion suggested by Nowick. That is, the measured activation energy is close to that for copper self-diffusion, the low temperature results are affected by grain-boundary diffusion, and the most reliable D value is the one obtained at the highest temperature. The latter point is supported by the agreement between these results and those of Matano at the higher temperatures, 970°C to 1000°C, even though the experimental methods were quite different.

Using Zener's[18] expression for D_0, Nowick has calculated a value of 0.23 cm²/sec for D_0 for the volume diffusion of gold into copper. Dienes[19] has proposed the following empirical relation between D_0 and Q,

$$D_0 = \nu\lambda^2 K e^{Q/RT_m}. \quad (2)$$

For the case of gold diffusing into copper, the vibrational frequency of copper, ν, is 6.6×10^{12} per second; the distance between nearest neighbors, λ, is 2.55×10^{-8} cm; the melting point T_m is 1357°K; and the empirical constant K equals 10^{-6}. R is the universal gas constant and Q is the activation heat of diffusion in the appropriate units. With the measured value of Q, 44 900 cal/mole, the calculated value of D_0 is 0.076 cm²/sec. This value is well within the probable error of the value, $D_0 = 0.10 \pm 0.06$ cm²/sec, found in this investigation. This empirical relation shows good agreement with several cases of self-diffusion and with many cases of intermetallic diffusion in which the concentration of the solute was low.

[12] B. V. Rollin, Phys. Rev. **55**, 231 (1939).
[13] Steigman, Shockley, and Nix, Phys. Rev. **56**, 13 (1939).
[14] Raynor, Thomassen, and Rouse, Trans. Am. Soc. Metals **30**, 313 (1942).
[15] M. S. Maier and H. R. Nelson, Trans. Am. Inst. Mining Met. Engrs. **147**, 39 (1942).
[16] G. Kuczynski, J. Metals **1**, 169 (1949).
[17] J. H. Dedrick and A. Gerds, J. Appl. Phys. **20**, 1042 (1949).
[18] C. Zener, J. Appl. Phys. **22**, 372 (1951).
[19] G. J. Dienes, J. Appl. Phys. **21**, 1191 (1950).

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 3 MARCH, 1954

General Theory of Plane-Wave Scattering from Finite, Conducting Obstacles with Application to the Two-Antenna Problems

J. E. Storer and J. Sevick
Cruft Laboratory, Harvard University, Cambridge, Massachusetts
(Received August 7, 1953)*

In this paper the variational techniques of Levine and Schwinger are used to obtain an approximate solution for the free-space scattering by a number of finite, perfectly conducting obstacles in which the mutual interactions or coupling are taken into account. The far-zone scattered field is found in terms of the obstacle currents. Variational principles for the total and backscattering cross sections are included. The special case of two parallel wires of finite length is worked out, and the results agree well with experiment.

PART I

1. Introduction

INVESTIGATORS have made little progress toward a solution of the old problem of the scattering of electromagnetic energy by obstacles in the path of a wave which includes their mutual interaction although Heaviside[1] pointed out this need long ago. Most authors neglect, as Lawson and Uhlenbeck[2] do in their discussion of "clutter," the excitation of one obstacle by another compared with its direct excitation. Reich[3] in 1916 obtained an iterated solution for a random distribution of Hertzian dipoles which included their interactions.

Much later Ament[4] and Theissing[5] solved approximately the case of a random distribution of spheres taking into account some of the interaction effects.

* The original unrevised manuscript was received in the editorial office on November 28, 1952.
[1] O. Heaviside, *Electromagnetic Theory* (Dover Publications, New York, 1950), Sec. 182.
[2] J. L. Lawson and G. E. Uhlenbeck, *Threshold Signals*, M. I. T. Radiation Laboratory Series (McGraw-Hill Book Company, Inc., New York, 1950).
[3] F. Reich, Ann. Physik **50**, 1 (1916); **50**, 121 (1916).
[4] W. S. Ament, "Preliminary Report on Electromagnetic Wave Propagation in a Composite Medium," NRL Report No. R-3238 (1948).
[5] H. H. Theissing, J. Opt. Soc. Am. **40**, 232 (1950).

Recently, Twersky[6] obtained a formal solution for the scattering of a plane electromagnetic wave by an arbitrary configuration of infinite parallel cylinders in which the sum of an infinite number of orders of scattered waves is added to the incident one. Twersky indicates over what ranges the single scattering hypotheses holds and where it breaks down. It is difficult to get numerical results where the cylinder radii or spacings are comparable with the incident wavelength, because Twersky's eigenvalue expansions then converge slowly.

In the present paper the variational techniques of Levine and Schwinger[7] are used to get approximate solutions for the three-dimensional electromagnetic scattering by a number of finite, perfectly conducting objects whose interactions are explicitly accounted for. The result gives the far zone scattered-field distribution in terms of the obstacle currents and has the usual variational properties, i.e., (1) it is stationary with respect to independent variations of the surface currents, and (2) it is independent of the scale of the surface currents. Variational principles are obtained for the total and the backscattering cross sections.

2. Integral Equation and Associated Variational Principle

Assuming a harmonic time dependence of the form $e^{-i\omega t}$, Maxwell's equations for the electromagnetic fields in free space are

$$\nabla\times\mathbf{H}(\mathbf{r})=-\frac{ik}{\zeta_0}\mathbf{E}(\mathbf{r}) \quad \nabla\times\mathbf{E}(\mathbf{r})=ik\zeta_0\mathbf{H}(\mathbf{r}), \tag{1}$$

where $k=\omega/c=2\pi/\lambda$, $\zeta_0=(u/\epsilon)^{\frac{1}{2}}$.

The problem we are interested in consists of a plane wave incident on a number of perfectly conducting obstacles. For this case it can be shown that[7,9]

$$\mathbf{E}(\mathbf{r})=\mathbf{E}^{\text{inc}}(\mathbf{r})+\mathbf{E}^{\text{sc}}(\mathbf{r}), \tag{2}$$

where $\mathbf{E}^{\text{inc}}(\mathbf{r})$ is the incident plane wave and $\mathbf{E}^{\text{sc}}(\mathbf{r})$ are the fields scattered by the obstacles. In particular

$$\mathbf{E}^{\text{sc}}(\mathbf{r})=\int_{\sigma_0}\Gamma(\mathbf{r},\mathbf{r}')\cdot\mathbf{K}(\mathbf{r}')d\sigma_0', \tag{3}$$

where

(1) $\Gamma(\mathbf{r},\mathbf{r}')$, the free-space dyadic Green's function, is given by

$$(\mathbf{r},\mathbf{r}')=ik\zeta_0\left[\epsilon-\frac{1}{k^2}\nabla\nabla'\right]\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}; \tag{4}$$

(2) $\mathbf{K}(\mathbf{r})$ is surface current flowing on the obstacles;

(3) σ_0 denotes collectively all the obstacle surfaces. [Thus, $\int_{\sigma_0}$ denotes an integration over each obstacle surface.]

Equation (3), thus, is just a statement of the fact that the scattered fields are produced by the currents flowing on the obstacles.

Since the obstacles are assumed to be perfectly conducting, $[\mathbf{n}\times\mathbf{E}(\mathbf{r})]=0$ on the surface of the obstacles. If we apply this to the representation of the electric field, (2) and (3), yield the set of vector integral equations

$$\mathbf{n}\times\left[\mathbf{E}^{\text{inc}}(\mathbf{r})+\int_{\sigma_0}\Gamma(\mathbf{r},\mathbf{r}')\mathbf{K}(\mathbf{r}')d\sigma_0'\right]_{\text{r on }\sigma_0}=0. \tag{5}$$

More specifically, if N obstacles are considered, Eq. (5) can be written as

$$[\mathbf{E}^{\text{inc}}(\mathbf{r})]_{\text{r on }\sigma_n}=-\sum_{m=1}^{N}\int_{\sigma_m}[\Gamma(\mathbf{r},\mathbf{r}')]_{\text{r on }\sigma_m}\cdot\mathbf{K}_m(\mathbf{r}')d\sigma', \quad n=1,2,\cdots N, \tag{6}$$

where σ_n denotes the surface of the nth obstacle and $\mathbf{K}_m(r')$ the surface current on the mth obstacle. Equation (6) forms a set of simultaneous integral equations for the surface current. If solutions to these integral equations could be found, then they could be substituted back in Eq. (3) and the scattered field determined. King[10] and Tai[11] employed a similar technique in finding the current distribution and impedance characteristics of coupled driven antennas. By making use of the symmetry present in their problem, they were able to reduce their two simultaneous integral equations to a pair of single integral equations that could be solved. However, this simplification due to symmetry is not obtained unless all obstacles are identical.

If the incident wave, which is assumed to be a plane wave of unit amplitude, is written as

$$\mathbf{E}_{\alpha,\beta}{}^{\text{inc}}(\mathbf{r})=\boldsymbol{\alpha}\exp(ik\boldsymbol{\beta}\cdot\mathbf{r}), \tag{7}$$

where $\boldsymbol{\alpha}$ is a unit vector in the direction of polarization, $\boldsymbol{\beta}$ is a unit vector in the direction of propagation, Eq. (5) can be written more explicitly in the form

$$\mathbf{E}_{\alpha,\beta}{}^{\text{inc}}(\mathbf{r})\Big|_{\text{r on }\sigma_0}=-\int_{\sigma_0}\Gamma(\mathbf{r},\mathbf{r}')\Big|_{\text{r on }\sigma_0}\cdot\mathbf{K}_{\alpha,\beta}(\mathbf{r}')d\sigma'. \tag{8}$$

The notation $\mathbf{K}_{\alpha,\beta}(\mathbf{r}')$ is understood to be the surface current which is induced by an incident plane electromagnetic wave characterized by $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

Consider a quantity defined by

$$A(\alpha,\beta;\alpha',\beta')=\int_{\sigma_0}\mathbf{E}_{\alpha,\beta}{}^{\text{inc}}(\mathbf{r})\mathbf{K}_{\alpha',\beta'}(\mathbf{r})d\sigma'. \tag{9}$$

[6] V. Twersky, "Multiple scattering of radiation, part I," Research Report No. Em-34, New York University, New York (July, 1951).

[7] H. Levine and J. Schwinger, "On the Theory of Electromagnetic Wave Diffraction by an Aperture in an Infinite Plane Conducting Screen," *Communications on Pure and Applied Mathematics* (Interscience Publishers, Inc., New York, 1950), Vol. III, No. 4.

[8] J. Schwinger, "The theory of obstacles in resonant cavities and wave guides," Radiation Laboratory Report 43-34, Research Laboratory of Electronics, M. I. T. (May 21, 1943).

[9] J. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), p. 464.

[10] R. King and C. W. Harrison, Jr., J. Appl. Phys. **15**, 481 (1944).

[11] C. T. Tai, "Theory of coupled antennas and its applications," Doctoral thesis, Harvard University (1946).

The quantity $A(\alpha,\beta;\alpha',\beta')$ is actually a surface integral, taken over the obstacles, of the scalar product of an incident electric field characterized by α and β and a surface current induced by an incident field characterized by α' and β'. Since

$$\Gamma(\mathbf{r}',\mathbf{r})=\Gamma(\mathbf{r},\mathbf{r}').$$

A substitution of Eq. (8) in (9) yields

$$\begin{aligned}A(\alpha,\beta;\alpha',\beta')&=\int_{\sigma_0}\int_{\sigma_0}[\mathbf{K}_{\alpha,\beta}(\mathbf{r})\cdot\Gamma(\mathbf{r},\mathbf{r}')]_{r\text{ on }\sigma_0}\cdot\mathbf{K}_{\alpha',\beta'}(\mathbf{r}')d\sigma'd\sigma\\&=\int_{\sigma_0}\mathbf{E}_{\alpha',\beta'}{}^{\text{inc}}(\mathbf{r})\cdot\mathbf{K}_{\alpha,\beta}(\mathbf{r})d\sigma\\&=A(\alpha',\beta';\alpha,\beta).\end{aligned}\tag{10}$$

The equality of $A(\alpha\beta,\alpha'\beta')$ and $A(\alpha'\beta',\alpha\beta)$ will later be shown to be a reciprocity relation for the excitation and observation along a pair of directions in space.

If we divide Eq. (10) by

$$\int_{\sigma_0}\mathbf{E}_{\alpha,\beta}{}^{\text{inc}}(\mathbf{r})\cdot\mathbf{K}_{\alpha',\beta'}(\mathbf{r})d\sigma\int_{\sigma_0}\mathbf{E}_{\alpha',\beta'}{}^{\text{inc}}(\mathbf{r})\cdot\mathbf{K}_{\alpha,\beta}(\mathbf{r})d\sigma$$

yields a bilateral variational principle in the form

$$\frac{1}{A(\alpha,\beta;\alpha',\beta')}=\frac{1}{A(\alpha',\beta';\alpha,\beta)}=\frac{-\int_0\int_{\sigma_0}\mathbf{K}_{\alpha,\beta}(\mathbf{r})\cdot\Gamma(\mathbf{r},\mathbf{r}')\Big|_{r\text{ on }\sigma_0}\cdot\mathbf{K}_{\alpha',\beta'}(\mathbf{r}')d\sigma'd\sigma}{\int_{\sigma_0}\mathbf{E}_{\alpha,\beta}{}^{\text{inc}}(\mathbf{r})\cdot\mathbf{K}_{\alpha',\beta'}(\mathbf{r})d\sigma\int_{\sigma_0}\mathbf{E}_{\alpha',\beta'}{}^{\text{inc}}(\mathbf{r})\cdot\mathbf{K}_{\alpha,\beta}(\mathbf{r})d\sigma}.\tag{11}$$

Equation (11) has all the usual variational properties. It can readily be shown that

$$\frac{\delta}{\delta K_{\alpha\beta}}A(\alpha,\beta;\alpha',\beta')=\frac{\delta}{\delta K_{\alpha'\beta'}}A(\alpha,\beta;\alpha',\beta')=0.$$

3. Relation between Variational Formulation and Scattered-Field Distribution

A scalar quantity, $A(\alpha,\beta;\alpha',\beta')$, was derived in Sec. 2 which possessed the usual variational properties. No real significance was given this quantity other than the fact that a sort of reciprocity existed between the incident field and a surface current which was induced by another field. In this section we will attempt to demonstrate the connection between this scalar quantity and the vector field scattered from finite obstacles.

In the far zone, where $r\gg r'$ (r' taken on obstacle), the dyadic Green's function is

$$\Gamma(\mathbf{r},\mathbf{r}')\simeq ik\zeta(\epsilon-\varrho\varrho)\frac{e^{ikr}}{4\pi r}\exp[-ik(\varrho\cdot\mathbf{r}')],\tag{12}$$

where $\varrho=\mathbf{r}/r$. This relation when inserted into the integral for the scattered field yields the far-zone scattered fields due to an incident field characterized by α and β. Thus,

$$\begin{aligned}\mathbf{E}_{\alpha\beta}{}^{\text{sc}}(\mathbf{r})&=\int_{\sigma_0}\Gamma(\mathbf{r},\mathbf{r}')K_{\alpha\beta}(\mathbf{r}')d\sigma'\\&=ik\zeta\frac{e^{ikr}}{4\pi r}\int_{\sigma_0}\exp[-ik(\varrho\cdot\mathbf{r}')](\epsilon-\varrho\varrho)\cdot\mathbf{K}_{\alpha\beta}(\mathbf{r}')d\sigma'\\&=ik\zeta\frac{e^{ikr}}{4\pi r}\mathbf{A}(\varrho;\alpha,\beta),\end{aligned}\tag{13}$$

where

$$\mathbf{A}(\varrho;\alpha,\beta)=\int_{\sigma_0}\exp[-ik(\mathbf{r}'\cdot\varrho)](\epsilon-\varrho\varrho)\cdot\mathbf{K}_{\alpha\beta}(r')d\sigma'.\tag{14}$$

The quantity $\mathbf{A}(\varrho,\alpha\beta)$ can be thought of as the far-zone distribution of the scattered field in the direction ϱ resulting from an induced current $\mathbf{K}_{\alpha\beta}(r')$. It is evident that the vector $\mathbf{A}(r,\alpha\beta)$ is perpendicular to the direction of propagation as

$$\varrho\cdot\mathbf{A}(r;\alpha,\beta)=\int_{\sigma_0}\exp[-ik(\mathbf{r}'\cdot\varrho)]\mathbf{r}\cdot(\epsilon-\varrho\varrho)\cdot\mathbf{K}_{\alpha\beta}(\mathbf{r}')d\sigma',$$

$$=0$$

since

$$\varrho\cdot(\epsilon-\varrho\varrho)=(\varrho-\varrho)=0.$$

Now if an arbitrary unit vector $\mathbf{s}$ is chosen such that it is perpendicular to the direction of propagation ϱ, i.e., $\mathbf{s}\cdot\varrho=0$, then

$$\begin{aligned}\mathbf{s}\cdot\mathbf{A}(\rho;\alpha,\beta)&=\int_{\sigma_0}\exp[-ik(\mathbf{r}'\cdot\varrho)]\mathbf{s}\cdot(\epsilon-\varrho\varrho)\cdot\mathbf{K}_{\alpha\beta}(\mathbf{r}')d\sigma'\\&=\int_{\sigma_0}\mathbf{s}\exp[-ik(\mathbf{r}'\cdot\varrho)]\mathbf{K}_{\alpha\beta}(\mathbf{r}')d\sigma',\end{aligned}\tag{15}$$

since

$$\mathbf{s}\cdot\epsilon=\mathbf{s}.$$

Since the notation used to describe the electric field is

$$\mathbf{E}_{\alpha,\beta}(\mathbf{r})=\alpha\exp[ik(\beta\cdot\mathbf{r})],$$

it is apparent that

$$\mathbf{s}\exp[-ik(\varrho\cdot\mathbf{r}')]=\mathbf{E}_{\mathbf{s},-\varrho}(\mathbf{r}').$$

Therefore, Eq. (15) can be rewritten as

$$\begin{aligned}\mathbf{s}\cdot\mathbf{A}(\varrho;\alpha,\beta)&=\int_{\sigma_0}\mathbf{E}_{s,-\rho}(\mathbf{r}')\cdot\mathbf{K}_{\alpha,\beta}(\mathbf{r}')d\sigma'\\&=A(\mathbf{s},-\varrho;\alpha,\beta).\end{aligned}\tag{16}$$

Thus, it is found that the scalar quantity $A(s,-\varrho;\alpha,\beta)$ is a component of the scattering distribution vector $\mathbf{A}(\varrho,\alpha\beta)$ along an arbitrary direction $\mathbf{s}$. From the result of Eq. (16) it is now possible to give a physical meaning to the statement of reciprocity made in Sec. 2. From Eq. (10) is was found that

$$A(\alpha,\beta;\alpha',\beta'')=A(\alpha',\beta';\alpha,\beta).$$

Hence, by Eq. (16) it follows that

$$A(\alpha,\beta;\alpha',\beta')=\boldsymbol{\alpha}\cdot\mathbf{A}(-\beta;\alpha',\beta'),$$
$$A(\alpha',\beta';\alpha,\beta)=\boldsymbol{\alpha}'\cdot\mathbf{A}(-\beta';\alpha,\beta).$$

Therefore,

$$\boldsymbol{\alpha}\cdot\mathbf{A}(-\beta;\alpha',\beta')=\boldsymbol{\alpha}'\cdot\mathbf{A}(\beta';\alpha,\beta). \tag{17}$$

Equation (17) states that a unit incident field characterized by $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is scattered with a component along $\boldsymbol{\alpha}'$ in the direction $-\boldsymbol{\beta}$ equal in magnitude to the component along $\boldsymbol{\alpha}$ in the direction $-\boldsymbol{\beta}$ due to a scattered unit incident field characterized by $\boldsymbol{\alpha}'$ and $\boldsymbol{\beta}'$. This reciprocity relation does not only take into account a pair of directions with respect to the incident polarization and the component of the scattered field.

4. A Theorem Concerning the Total Scattering Cross Section and the Scattered-Field Distribution

The total scattering cross section is defined as the ratio of the total scattered power to the incident power per unit area. Mathematically, it can be written as

$$\sigma=\frac{\tfrac{1}{2}\,\mathrm{Re}\int_{\sigma_0}\mathbf{E}^{\mathrm{sc}*}(\mathbf{r})\times\mathbf{H}^{\mathrm{sc}}(\mathbf{r})\cdot\mathbf{n}d\sigma}{\tfrac{1}{2}\,\mathrm{Re}\mathbf{E}^{\mathrm{inc}*}(\mathbf{r})\times\mathbf{H}^{\mathrm{inc}}(\mathbf{r})\cdot\boldsymbol{\beta}}. \tag{18}$$

It is understood in Eq. (18) that $\mathbf{n}$ is the outward normal to the scattering obstacles and $\boldsymbol{\beta}$ is the direction of propagation of the incident field. With the previous notation and the equality $\nabla\times\mathbf{E}=ik\zeta\mathbf{H}$, the denominator of (18) gives

$$\begin{aligned}
&\tfrac{1}{2}\,\mathrm{Re}\mathbf{E}_{\alpha\beta}{}^{\mathrm{inc}}(\mathbf{r})\times H_{\alpha\beta}{}^{\mathrm{inc}*}(\mathbf{r})\cdot\boldsymbol{\beta}\\
&=\tfrac{1}{2}\,\mathrm{Re}\{\boldsymbol{\alpha}\exp[ik(\boldsymbol{\beta}\cdot\mathbf{r})]\}^*\\
&\qquad\times\left\{\frac{1}{ik\zeta}\nabla\times\boldsymbol{\alpha}\exp[ik(\boldsymbol{\beta}\cdot\mathbf{r})]\right\}\cdot\boldsymbol{\beta}\\
&=\tfrac{1}{2}\,\mathrm{Re}\{\boldsymbol{\alpha}\exp[-ik(\boldsymbol{\beta}\cdot\mathbf{r})]\}\\
&\qquad\times\left\{-\frac{1}{ik\zeta}\boldsymbol{\alpha}\times\nabla\exp[ik(\boldsymbol{\beta}\cdot\mathbf{r})]\right\}\cdot\boldsymbol{\beta}\\
&=\tfrac{1}{2}\,\mathrm{Re}\frac{-\boldsymbol{\alpha}\times(\boldsymbol{\alpha}\times\boldsymbol{\beta})\cdot\boldsymbol{\beta}}{\zeta}\\
&=\frac{1}{2\zeta}.
\end{aligned} \tag{19}$$

At the surface of the obstacles $E_{\alpha\beta}{}^{\mathrm{sc}*}(r)=-E_{\alpha\beta}{}^{\mathrm{inc}*}(r)$; hence

$$\sigma=\mathrm{Re}\left[-\zeta\int_{\sigma}E_{\alpha\beta}{}^{\mathrm{inc}*}(\mathbf{r})\times H_{\alpha\beta}{}^{\mathrm{sc}}(\mathbf{r})\cdot\mathbf{n}d\sigma\right]. \tag{20}$$

Since the magnetic field can be written as

$$\mathbf{H}_{\alpha\beta}{}^{\mathrm{sc}}(\mathbf{r})=\mathbf{H}_{\alpha\beta}(\mathbf{r})-\mathbf{H}_{\alpha\beta}{}^{\mathrm{inc}}(\mathbf{r})$$

and since

$$\int\mathbf{E}_{\alpha\beta}{}^{\mathrm{inc}*}(\mathbf{r})\times\mathbf{H}_{\alpha\beta}{}^{\mathrm{inc}}(\mathbf{r})\cdot\mathbf{n}d\sigma$$

represents no transfer of power, it follows that

$$\begin{aligned}
\sigma&=\mathrm{Re}-\zeta\int_{\sigma_0}\mathbf{E}_{\alpha\beta}{}^{\mathrm{inc}*}(\mathbf{r})\times\mathbf{H}_{\alpha\beta}(\mathbf{r})\cdot\mathbf{n}d\sigma\\
&=\mathrm{Re}\zeta\int_{\sigma_0}\mathbf{E}_{\alpha\beta}{}^{\mathrm{inc}*}(\mathbf{r})\cdot(\mathbf{n}\times\mathbf{H}_{\alpha\beta}(\mathbf{r}))d\sigma\\
&=\zeta\,\mathrm{Re}\int_{\sigma}\mathbf{E}_{\alpha\beta}{}^{\mathrm{inc}*}(\mathbf{r})\cdot\mathbf{K}_{\alpha\beta}(\mathbf{r})d\sigma\\
&=\zeta\,\mathrm{Re}A(\alpha,-\beta;\alpha,\beta).
\end{aligned} \tag{21}$$

Hence a variational expression has been obtained for the total scattering cross section.

5. The Backscattering Cross Section

From the development by the preceding sections it is possible to derive, in general, another parameter which is directly useful in the radar problem. Since most radar applications are concerned with the information about the scattered power in the direction of the radar antenna, a more applicable term is the backscattering cross section. This term may be defined, in general, as the total-scattering cross section of a fictitious isotropic scatterer which scatters energy in all directions with intensity equal to that scattered directly back toward the source by the actual obstacle concerned. In the derivation of the radar equation, which is an expression for the power density of the scattered field at the radar antenna, several writers[12–14] consider the scattering obstacles as having (1) an effective area that intercepts the incident flux and scatters it isotropically and (2) a gain or reflection factor. The product of quantities (1) and (2) is a measure of the backscattering ability. It is more common, however, to lump the product of this effective area and gain factor into a single quantity, σ_t, known as the "backscattering cross section." Mathematically it is

[12] M. I. T. Radar School Staff, *Principles of Radar* (McGraw-Hill Book Company, Inc., New York, 1946), Chap. I, Art. 8.

[13] L. Ridenour, editor, *Radar System Engineering*, written by E. Purcell (McGraw-Hill Book Company, Inc., New York, 1947), Chap. 2.

[14] S. Silver, *Microwave Antenna Theory and Design*, M. I. T. Radiation Laboratory Series (McGraw-Hill Book Company, Inc., New York, 1949), Vol. 12, p. 468.

defined as[15]

$$\sigma_t = 4\pi R_0^2 \left| \frac{\mathbf{S}_{R_0}^{sc}}{\mathbf{S}^{inc}} \right|, \tag{22}$$

where $\mathbf{S}_{R_0}^{sc}$ is the "backscattered" Poynting vector at radius R_0 from the center of the scatterer, and $\mathbf{S}^{inc}$ is the "incident" Poynting vector at the center of the scatterer. More explicitly, the magnitude of the Poynting vectors are

$$|\mathbf{S}_{R_0}| = \tfrac{1}{2} \operatorname{Re}(\mathbf{E}_{\alpha\beta}^{sc*}(r) \times \mathbf{H}_{\alpha\beta}^{sc}(r) \cdot (-\beta)|_{r \text{ on } R_0}) \tag{23}$$

$$|\mathbf{S}^{inc}| = \tfrac{1}{2} \operatorname{Re}(\mathbf{E}_{\alpha\beta}^{inc*}(r) \times \mathbf{H}_{\alpha\beta}^{inc}(r) \cdot \beta). \tag{24}$$

From Eq. (19), (24) reduces to

$$\tfrac{1}{2} \operatorname{Re}(\mathbf{E}_{\alpha\beta}^{*inc}(\mathbf{r}) \times \mathbf{H}_{\alpha\beta}^{inc}(\mathbf{r}) \cdot \beta) = \frac{1}{2\zeta}.$$

From the derivations given in Sec. 3, the backscattered field caused by an incident field characterized by α and β can be written as

$$\mathbf{E}_{\alpha\beta}^{sc}(r) = ik\zeta \frac{e^{ikr}}{4\pi r} \mathbf{A}(-\beta,\alpha\beta).$$

Thus the complex conjugate of the scattered field is

$$\mathbf{E}_{\alpha\beta}^{sc*}(r) = -ik\zeta \frac{e^{-ikr}}{4\pi r} \mathbf{A}^*(-\beta,\alpha\beta).$$

Therefore, Eq. (22) results in

$$\begin{aligned}
\sigma_t &= 4\pi R_0^2 \zeta \operatorname{Re}\left[-ik\zeta \frac{e^{-ikr}}{4\pi r} \mathbf{A}^*(-\beta,\alpha\beta) \right. \\
&\quad \left. \times \left(\frac{1}{ik\zeta} \nabla \times \frac{ik\zeta e^{ikr}}{4\pi r} \mathbf{A}(-\beta,\alpha\beta) \right) \cdot (-\beta) \right]_{r \text{ on } R_0} \\
&= 4\pi R_0^2 \zeta \operatorname{Re}\left[-ik\zeta \frac{e^{-ikr}}{4\pi r} \mathbf{A}^*(-\beta,\alpha\beta) \right. \\
&\quad \left. \times \left(-ik\beta \times \mathbf{A}(-\beta,\alpha\beta) \frac{e^{ikr}}{4\pi r} \right) \cdot (-\beta) \right]_{r \text{ on } R_0} \\
&= \frac{\zeta^2 k^2}{4\pi} \operatorname{Re}[\mathbf{A}^*(-\beta,\alpha\beta) \times (\beta \times \mathbf{A}(-\beta,\alpha\beta)) \cdot \beta] \\
&= \frac{\zeta^2 k^2}{4\pi} |\mathbf{A}(-\beta,\alpha\beta)|^2.
\end{aligned} \tag{25}$$

The quantity $\mathbf{A}(-\beta,\alpha\beta)$ can be broken down into two orthogonal components which form a plane perpendicular to the direction of propagation $-\beta$. One component can be chosen along α, the polarization of the incident field, and the other perpendicular to α. Thus it can be written that

$$\mathbf{A}(-\beta,\alpha\beta) = \alpha A(\alpha,\beta;\alpha,\beta) + \alpha \times \beta A(\alpha \times \beta, \beta; \alpha,\beta). \tag{26}$$

Therefore the substitution of Eq. (26) in Eq. (25) gives, for the backscattering cross section,

$$\sigma_t = \frac{\zeta^2 k^2}{4\pi} (|A(\alpha,\beta;\alpha,\beta)|^2 + |A(\alpha \times \beta, \beta; \alpha,\beta)|^2). \tag{27}$$

Hence an expression for the backscattering cross section has been obtained in terms of stationary quantities.

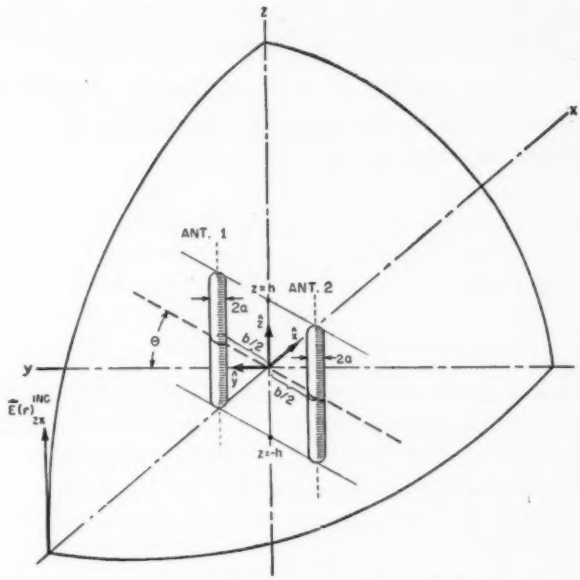


FIG. 1. Orientation of two identical unloaded antennas.

PART II

1. Introduction

The general results of Part I lend themselves very readily to the study of the backscattering cross section of two identical, nonstaggered, unloaded antennas which are oriented parallel to the incident field as shown in Fig. 1. This particular problem has the following advantages:

1. It renders trial functions (not including the region of "long antennas") which are known and readily integrable.
2. It gives large scattering, thereby facilitating experimental verification.
3. It offers an opportunity for comparison with the great amount of work previously done in the antenna field.

This part of the report will present the complete mathematical development of the problem and a final curve for the case of broadside radiation ($\theta = 0°$) for the half-wave and full-wave antennas. Other cases consisting of varying thicknesses, heights, and angles of θ will be treated in the succeeding papers.

[15] R. King, *Notes on Antennas* (Cruft Laboratory, 1949) (to be published as *Theory of Linear Antennas* by Harvard University Press, Cambridge).

2. Reduction of the General Formulation of the Backscattering Cross Section to Two Antennas

For the analysis of the two-antenna problem as shown in Fig. 1, the following assumptions are added to that of perfect conductivity:

$$k^2a^2 \ll 1, \qquad a^2/h^2 \ll 1, \qquad a^2/b^2 \ll 1, \tag{28}$$

where a is the radius of antennas, h is the half-length, and b is the spacing between centers. With the scattering antennas oriented along the direction of polarization of the incident field, it is found from the general formulation that only one component of the scattered field is maintained, namely,

$$A(\alpha\times\beta,\beta;\alpha,\beta)=\int_{\sigma_0}\mathbf{E}_{\alpha\times\beta,\beta}{}^{\text{inc}}(r')K_{\alpha\beta}(r')d\sigma' =0,$$

since $\mathbf{E}_{\alpha\times\beta,\beta}{}^{\text{inc}}(r)$ does not exist. The backscattering cross section for the two thin antennas in free space can now be written as

$$\sigma_t=\frac{\zeta_0{}^2k^2}{4\pi}|A(\alpha,\beta;\alpha,\beta)|^2. \tag{29}$$

If we write the surface integral for the component as

$$\int_{\sigma_0}=\int_{\sigma_1}+\int_{\sigma_2}$$

and $\mathbf{E}_{\alpha,\beta}{}^{\text{inc}}(r)=\mathbf{z}\exp(ik\mathbf{x}\cdot\mathbf{r})$, the component in Eq. (29) becomes

$$A(\alpha,\beta;\alpha,\beta)=A(z,x;z,x)=\frac{\left\{\left(\int_{\sigma_1}+\int_{\sigma_2}\right)\mathbf{z}\exp[ik(\mathbf{x}\cdot\mathbf{r})]\mathbf{K}_{z,z}(\mathbf{r})d\sigma\right\}^2}{-\left(\int_{\sigma_1}+\int_{\sigma_2}\right)\left(\int_{\sigma_1}+\int_{\sigma_2}\right)\mathbf{K}_{z,z}(\mathbf{r})\cdot\Gamma(\mathbf{r},\mathbf{r}')\cdot\mathbf{K}_{z,z}(\mathbf{r}')d\sigma d\sigma'}.$$

With the first assumption of Eq. (28) and the orientation of Fig. 1, the surface current has only a z component, and hence $\mathbf{K}_{z,z}(\mathbf{z})$ can be written

$$\mathbf{K}(\mathbf{z})=\frac{I(z)}{2\pi a}\mathbf{z},$$

where $I(z)$ is the total current. This reduces the above equation to

$$A(z,x;z,x)=\frac{\frac{1}{2\pi a}\left[\left(\int_{\sigma_1}+\int_{\sigma_2}\right)\exp(ik\mathbf{x}\cdot\mathbf{r})I(z)d\sigma\right]^2}{\frac{-ik\zeta_0}{4\pi^2a^2}\left[\left(\int_{\sigma_1}+\int_{\sigma_2}\right)\left(\int_{\sigma_1}+\int_{\sigma_2}\right)I(z)I(z')G(r,r')d\sigma d\sigma'\right]}. \tag{30}$$

The form of the integrand in the denominator of Eq. (30) is explicitly

$$G(r,r')=\frac{1}{ik\zeta}(\mathbf{z}\cdot\Gamma(r,r')\cdot\mathbf{z})=\left(1+\frac{1}{k^2}\frac{\partial^2}{\partial z^2}\right)\frac{\exp(ik|\mathbf{r}-\mathbf{r}'|)}{4\pi|\mathbf{r}-\mathbf{r}'|}.$$

It is quite apparent that the remaining procedure is to select suitable trial functions for the currents, insert them into Eq. (30), and calculate values of the backscattering cross section. However, since the problem concerns two obstacles, the variational principle can be applied to Eq. (30) in order to find an optimum relation existing between the two currents. This procedure is similar to finding higher-order solutions for variational principles derived from a single-integral equation. In order to carry out this process, the method of resolving the currents into symmetrical and antisymmetrical components is used.

King and Harrison[16] have shown that for two identical antennas the currents I_1 and I_2 can be resolved into a codirectional or symmetrical pair and an oppositely directed or antisymmetrical pair. Whatever the magnitudes and relative phases of I_1 and I_2, they can always be expressed as follows:

$$I_1=I_s+I_a$$
$$I_2=I_s-I_a.$$

The problem now is to select satisfactory distribution functions for the symmetrical and antisymmetrical components. Considerable aid is given here by the recent work of Moritz[17] who has shown both theo-

[16] *Tables of Generalized Sine and Cosine Integral Function*, Annals of the Computation Laboratory of Harvard University, Vol. XVIII, Cambridge (1949).

[17] The reader is referred to the original report, T. R. 149, "A general theory of plane-wave scattering from finite conducting obstacles with application to the two-antenna problem," J. Sevick and J. E. Storer, Cruft Laboratory (April 10, 1952).

retically and experimentally that, for two identical, unloaded half-wavelength antennas, the symmetrical and antisymmetrical current distributions are well approximated by a shifted cosine. This held true for the antisymmetrical distribution for full-wavelength antennas while the symmetrical distribution differed only by the addition of a small hump at the ends of the antennas. (Thus, if the scattering antennas are restricted to lengths of a wavelength or less, a suitable trial function for either component is a shifted cosine.) It should be noted that this restriction is not a severe one since a large portion of the antenna work has been done in this region. Therefore, the currents can be written as

$$I_1(z)=(I_s+I_a)(\cos kh-\cos kz)$$
$$I_2(z)=(I_s-I_a)(\cos kh-\cos kz). \tag{31}$$

If we use trial functions (31) and note with the use of Fig. 2 that, to order k^2a^2, $\exp(ik\mathbf{x}\cdot\mathbf{r})$ can be replaced in the integral with $\exp(ikx_1)$ for antenna 1 and $\exp(-ikx_1)$ for antenna 2, Eq. (30) becomes

$$A(z,x;z,x)=\frac{\left[\frac{1}{2\pi a}e^{ikx_1}(I_a+I_s)\int_{\sigma_1}f(z)d\sigma+\frac{1}{2\pi a}e^{-ikx_1}(I_a-I_s)\int_{\sigma_2}f(z)d\sigma\right]^2}{-\frac{ik\zeta_0}{4\pi^2a^2}\left\{(I_a+I_s)^2\int_{\sigma_1}\int_{\sigma_1}f(z)f(z')G(r,r')d\sigma d\sigma'+2(I_a+I_s)(I_a-I_s)\int_{\sigma_1}\int_{\sigma_2}f(z)f(z')G(r,r')d\sigma d\sigma'\right.}$$
$$\left.+(I_a-I_s)^2\int_{\sigma_2}\int_{\sigma_2}f(z)f(z')G(r,r')d\sigma d\sigma'\right\}, \tag{32}$$

where $f(z)=\cos kh-\cos kz'$. Since the antennas are identical, Eq. (32) can be reduced to

$$A(z,x;z,x)=\frac{\gamma^2[(I_s+I_a)e^{ikx_1}+(I_s-I_a)e^{-ikx_1}]^2}{-(ik\zeta_0/4\pi^2)[2\gamma_1(I_s^2+I_a^2)+2\gamma_2(I_s^2-I_a^2)]}$$
$$=\frac{i8\pi^2\gamma^2}{k\zeta_0}\frac{[I_s\cos kx_1+iI_a\sin kx_1]^2}{[\gamma_1(I_s^2+I_a^2)+\gamma_2(I_s^2-I_a^2)]}, \tag{33}$$

where

$$\gamma=\int_{-h}^{h}f(z)dz=\frac{2}{k}(kh\cos kh-\sin kh)$$

$$\gamma_1,\gamma_2=\int_{-h}^{h}\int_{-h}^{h}\int_{0}^{2\pi}\int_{0}^{2\pi}f(z)f(z')\times G_{1,2}(z-z',\phi-\phi')d\phi d\phi' dz dz'.$$

The integral γ_1 is taken over a single antenna, while γ_2 is taken over both antennas. Each integral, in turn, determines the form of the kernel $G_{1,2}(z-z'\ \phi-\phi')$. These γ_1, γ_2 integrals, can be evaluated in terms of the generalized sine and cosine integrals.[17]

According to Eq. (33), $A(z,x;z,x)$ is a function of two parameters I_s and I_a. Since the variational principle, as shown in Part I, is homogeneous in the currents I_1 and I_2, only one parameter must exist if the magnitudes of the currents are written as the sum of the difference of two arbitrary constants. Thus, one of these parameters can be eliminated by letting

$$\eta=I_a/I_s. \tag{34}$$

In carrying out the substitution in Eq. (33), the result yields

$$A(z,x;z,x)=\frac{i8\pi^2\gamma^2}{k\zeta_0}\frac{[\cos kx_1+\eta\sin kx_1]^2}{[(\gamma_1+\gamma_2)+(\gamma_1-\gamma_2)\eta^2]}. \tag{35}$$

In order to find the optimum value of η, the variationa procedure is again applied; that is, $A(z,x;z,x)$ is differentiated with respect to η, the result is equated to zero, and then solved for η. The procedure is as follows:

$$\frac{\partial A(z,x;z,x)}{\partial\eta}=\frac{4\gamma^2[\cos kx_1+i\eta\sin kx_1]i\sin kx_1}{(\gamma_1+\gamma_2)+(\gamma_1-\gamma_s)\eta^2}$$
$$-\frac{4\gamma^2[\cos kx_1+i\eta\sin kx_1]^2(\gamma_1-\gamma_2)\eta}{[(\gamma_1+\gamma_2)+(\gamma_1-\gamma_2)\eta^2]}$$
$$=0. \tag{36}$$

If we let

$$\alpha_1=\cos kx_1$$
$$\alpha_2=i\sin kx_1,$$

Eq. (36) gives

$$\alpha_2[(\gamma_1+\gamma_2)+(\gamma_1-\gamma_2)\eta^2]-[\alpha_1+\eta\alpha_2](\gamma_1-\gamma_2)\eta=0$$

and

$$\eta=\frac{\alpha_2(\gamma_1+\gamma_2)}{\alpha_1(\gamma_1-\gamma_2)}.$$

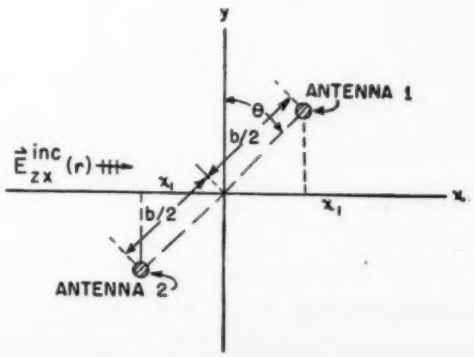


FIG. 2. Coordinates used for two antennas.

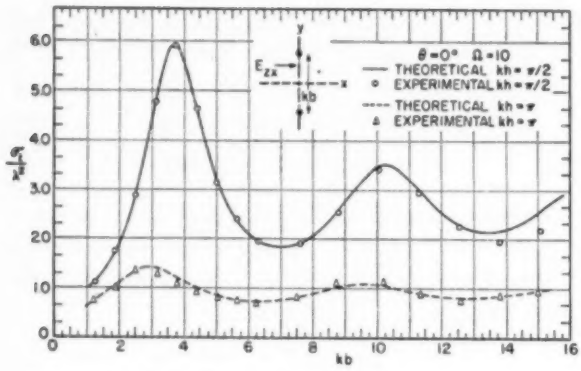


FIG. 3. Backscattering from two antennas.

Substitution of this into Eq. (35) gives, for the component $A(z,x;z,x)$,

$$A(z,x;z,x)=\frac{i8\pi^2\gamma^2}{k\zeta_0}\frac{\left[\alpha_1+\frac{\alpha_2^2}{\alpha_1}\frac{(\gamma_1+\gamma_2)}{(\gamma_1-\gamma_2)}\right]^2}{\left[(\gamma_1+\gamma_2)+(\gamma_1-\gamma_2)\frac{\alpha_2^2}{\alpha_1^2}\frac{(\gamma_1+\gamma_2)^2}{(\gamma_1-\gamma_2)^2}\right]}$$

$$=\frac{i8\pi^2\gamma^2}{k\zeta_0}\cdot\frac{\alpha_1}{[\gamma_1+\gamma_2]}\cdot\left[\alpha_1+\frac{\alpha_2^2}{\alpha_1}\frac{(\gamma_1+\gamma_2)}{(\gamma_1-\gamma_2)}\right]$$

$$=\frac{i8\pi^2\gamma^2}{k\zeta_0}\left[\frac{\alpha_1^2(\gamma_1-\gamma_2)+\alpha_2^2(\gamma_1+\gamma_2)}{(\gamma_1^2-\gamma_2^2)}\right].$$

Since $\alpha_1=\cos kx_1$, $\alpha_2=i\sin kx_1$, this reduces to

$$A(z,x;z,x)=\frac{i8\pi^2\gamma^2}{k\zeta_0}\left[\frac{\cos^2kx_1(\gamma_1-\gamma_2)-\sin^2kx_1(\gamma_1+\gamma_2)}{(\gamma_1^2-\gamma_2^2)}\right]$$

$$=\frac{i8\pi^2\gamma^2}{k\zeta_0}\left[\frac{\gamma_1\cos 2kx_1-\gamma_2}{\gamma_1^2-\gamma_2^2}\right]. \tag{37}$$

With the result of Eq. (37) inserted in Eq. (29), the final general form for the backscattering cross section of two identical nonstaggered antennas becomes

$$\sigma_t=16\pi^2\gamma^4\left[\frac{\gamma_1\cos 2kx_1-\gamma_2}{\gamma_1^2-\gamma_2^2}\right]^2. \tag{38}$$

The evaluation of the γ integrals yields[17]

$$\gamma=\frac{2}{k}(kh\cos kh-\sin kh),$$

$$\gamma_1=\frac{2\pi}{k}\Big[(2kh\cos^2kh-\sin 2kh)$$

$$\times\left\{-C(0,2kh)+iS(0,2kh)+\ln\frac{2h}{a}+\ln 2\right\}$$

$$-\cos^2kh\left\{\sin 2kh-i(\cos 2kh-1)\right\}$$

$$+C_s(0,2kh)+iS_s(0,2kh)+\frac{4ak}{\pi}(\cos^2kh-1)\Big],$$

$$\gamma_2=\frac{2\pi}{k}\Big[(2kh\cos^2kh-\sin 2kh)$$

$$\times\left\{\sinh^{-1}\frac{2kh}{kb}-C(kb,2kh)+iS(kb,2kh)\right\}$$

$$-\cos^2kh\{\sin(4k^2h^2+k^2b^2)^{\frac{1}{2}}-\sin kb$$

$$-i\cos(4k^2h^2+k^2b^2)^{\frac{1}{2}}+i\cos kb\}$$

$$+C_s(kb,2kh)+iS_s(kb,2kh)\Big].$$

The results for the broadside case, $\theta=0°$, for the half-wave and full-wavelength antennas are shown in Fig. 3. In this problem the radii and heights of the antennas are held fixed while the spacing between them is varied. It is noticed from Fig. 3 that the two curves, although not differing appreciably in shape, have very large differences in magnitude. This is explained by the fact that the resonant current of a full-wavelength antenna is an odd function and hence cannot be excited by an incident plane wave polarized parallel to the antenna. Also readily seen is the effect of mutual coupling. It is quite apparent from Fig. 3 that the "single-scattering" hypothesis is not a good approximation for antennas with spacing even as large as two wavelengths.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 3 MARCH, 1954

On a Nonlinear Diffusion Equation Applied to the Magnetization of Saturable Reactors

SHOU-HSIEN CHOW
Research Center, Burroughs Corporation, Philadelphia, Pennsylvania
(Received August 10, 1953)

The penetration of the magnetic field intensity into the interior of an infinite slab of ferromagnetic material having a nonlinear $B-H$ relationship is governed by a nonlinear diffusion equation. For finite thickness of the slab, the problem is solved by an approximate method. A rigorous solution of the same problem when the thickness of the slab approaches infinity is given to justify the approximate method used. In the paper, one term only is used to represent an infinite series in order to give a readily calculable result in "switching time." The result is then compared with that of the graphical solution of the same problem.

THE PROBLEM

THE problem treated here is the penetration of the magnetic field intensity H into an infinite slab of ferromagnetic material of thickness δ when a magnetic field intensity H_0 in the z direction is applied to the boundaries of the slab at $x=0$ and $x=\delta$ at $t=0$. The $B-H$ relationship of this material is that shown in Fig. 1 and the conductivity of the material is σ. The initial condition is $H=0$.

This represents closely the condition of a core made of thin laminations of metallic ferromagnetic material having a rectangular hysteresis loop, and excited by a magnetomotive force supplied by a constant current source. This problem is important in magnetic memories, magnetic shift register, and pulse transformers.

Maxwell's equations which lead to the diffusion equation read

$$\frac{\partial H_z}{\partial x}=-\sigma E_y, \tag{1}$$

$$\frac{\partial E_y}{\partial x}=-\frac{\partial B_z}{\partial t}, \tag{2}$$

and the $B-H$ relationship is given by

$$\begin{aligned} &B=\mu H \text{ when } 0<H<H_s \text{ (or unsaturated)},\\ &B=B_s=\mu H_s \text{ when } H>H_s \text{ (or saturated)}. \end{aligned} \tag{3}$$

On combining (1) and (2) and omitting subscripts, it is found that

$$\frac{\partial^2 H}{\partial x^2}=\sigma\frac{\partial B}{\partial t}. \tag{4}$$

When $H_0>H_s$, Eq. (3) introduces nonlinearity into the diffusion equation. The ratio H_s/H_0 will be denoted by ρ $(0\leq\rho\leq 1)$.

When δ approaches infinity, the problem can be solved rigorously (see Appendix). For finite values of δ, the problem will be solved by an approximate method.

PHYSICS OF THE PROBLEM

The diffusion of the magnetic field intensity into the interior of the slab can best be explained by plotting the $H-x$ distribution at different instants of time. The graphical solution of this problem was formulated by D. F. Hunt.[1] For values of ρ equal to 0.1, 0.3, 0.5, 0.7, 0.9, solutions have been obtained by the author. The so called "switching time" T_s, or the time when the magnetic field intensity at the center of the slab $(x=\delta/2)$ reaches H_s, is plotted against ρ in Fig. 2.

The essence of the graphical method is that the $H-x$ curve at a certain instant has a straight line segment between $x=0$ and $x=x_s$, where x_s is the coordinate at which $H=H_s$, $\partial H/\partial x$ is continuous at x_s, and H satisfies the relationship governed by the linear diffusion equation when $H<H_s$. The linear diffusion equation has been discussed in many books.[2]

Graphical solutions for $\rho=0.8$ and $\rho=0.5$ are reproduced here with only a small number of $H-x$ curves shown (Figs. 3 and 4). The graphical solutions indicate that when ρ approaches 0, $H-x$ curves become straight lines perpendicular to the x axis. When this is true, the diffusion is solely accomplished by the moving of the "saturation distance" x_s at a velocity dx_s/dt. In this case dx_s/dt can be easily found in the following manner.

In the saturated region between $x=0$ and $x=x_s$ (see Fig. 3),

$$\frac{\partial E}{\partial x}=0, \tag{5}$$

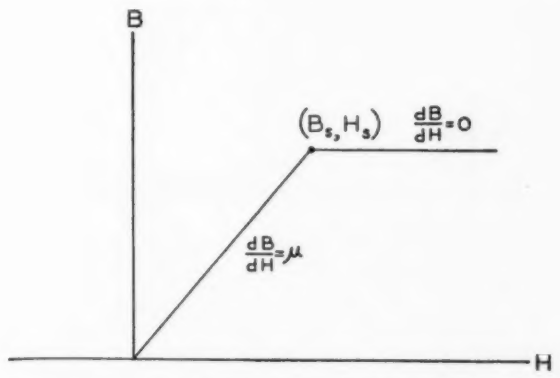


FIG. 1. $B-H$ characteristic.

[1] D. F. Hunt, *Saturation Time for Deltamax Transformer*, 1951, Internal Report, Burroughs Corporation.
[2] See, for instance, H. S. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids* (Clarendon Press, Oxford, 1947), Chap. III.

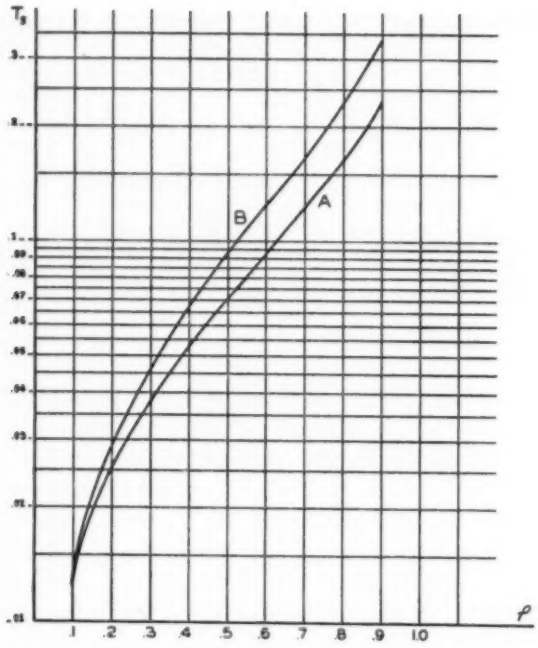


FIG. 2. Switching time T_s (in the unit of $\mu\sigma\delta^2$) for different values of ρ. (A) Graphical result. (B) Analytical result.

or

$$E=E_s=\frac{H_0-H_s}{\sigma x_s}. \tag{6}$$

This electric field intensity, acting for a time interval dt will cause a flux change per unit height of the slab given by the following equation:

$$E_s dt=\mu H_s dx_s=B_s dx_s. \tag{7}$$

On combining (6) and (7), it is found that

$$x_s=\left[\frac{2(H_0-H_s)t}{\sigma B_s}\right]^{\frac{1}{2}}, \tag{8}$$

and

$$E_s=\left[\frac{B_s(H_0-H_s)}{2\sigma t}\right]^{\frac{1}{2}}. \tag{9}$$

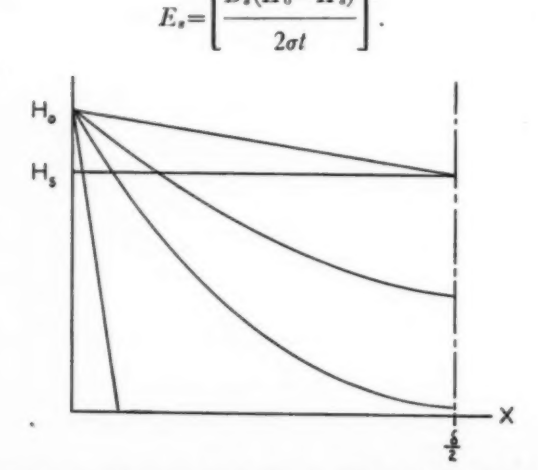


FIG. 3. $H-x$ curves at different instants of time ($\rho=0.8$).

When $x_s=\delta/2$, the penetration is complete, and the switching time is

$$T_s=\frac{\mu\sigma\delta^2}{8}\frac{H_s}{H_0-H_s}. \tag{10}$$

If k is defined as

$$k=\frac{E_s}{B_s(dx_s/dt)}, \tag{11}$$

it approaches unity when ρ approaches zero.

SOLUTION OF THE PROBLEM

When ρ is comparable to unity, the above approximation does not hold true because both the deviation of $H-x$ curves from straight lines and the magnetic field intensity at $x=\delta/2$ are not negligible. These two factors have opposite effects on the moving of x_s. The first one will tend to decrease dx_s/dt or will tend to make $k>1$. This can be seen from the fact that the separation of $H-x$ curves at two instants of time for the same value of H is a monotonically decreasing function of H (Fig. 3). The second factor will tend to increase dx_s/dt.

The magnetic field intensity $H(x,t)$, in this case, may be expressed as

$$H(x,t)=H(x,E_s,t)=H(x,x_s,t),$$

and therefore,

$$dH=\frac{\partial H}{\partial t}dt+\frac{\partial H}{\partial x_s}dx_s+\frac{\partial H}{\partial x}dx. \tag{12}$$

When the $H-x$ curves are represented by straight lines, it means

$$\frac{\partial^2 H}{\partial x^2}=\sigma\frac{\partial B}{\partial t}=0. \tag{13}$$

In the formulation which follows, $\partial B/\partial t$ or $\partial H/\partial t$ in the unsaturated region will be considered as a perturbation term. The total change in flux per unit height of the slab during Δt represented by the area *abcdfha* in Fig. 5 is the sum of the area of the trapezoid *abega* and a small area which is the difference between the areas *bcdeb* and *ahfga*. Expressions for the $H-x$ curves *bcd* and *ahf* may be found in the following manner.

Consider the following linear diffusion problem. A magnetic field intensity $H=H_s$ is applied to the boundaries of a previously unmagnetized slab at $x=x_s$ and $x=\delta-x_s$ at $t'=0$. For convenience, the x axis will be translated by x_s. On the new x' axis, the boundaries become $x'=0$ and $x'=\delta-2x_s=2a$. The linear diffusion equation $\partial^2 H(x',t')/\partial x'^2=\mu\sigma\partial H(x',t')/\partial t'$ can be solved by the method of Laplace transformation. The transformed equation reads

$$\frac{d^2 h(x',p)}{dx'^2}=\mu\sigma p h(x',p), \tag{14}$$

where $h(x',p)$ is the Laplace transform of $H(x',t')$. The solution of Eq. (14) is given by

$$h(x',p)=A\exp[(\mu\sigma p)^{\frac{1}{2}}x']+B\exp[-(\mu\sigma p)^{\frac{1}{2}}x']. \quad (15)$$

The constants A and B are then evaluated from the boundary conditions $H(0,t')=H(2a,t')=H_s$, or $h(0,p)=h(2a,p)=H_s/p$. Equation (15) becomes

$$h(x',p)=\frac{H_s}{p}\{\exp[(\mu\sigma p)^{\frac{1}{2}}(a-x')]-\exp[-(\mu\sigma p)^{\frac{1}{2}}(a-x')]+\exp[(\mu\sigma p)^{\frac{1}{2}}x']-\exp[(-\mu\sigma p)^{\frac{1}{2}}x']\}\times\sum_{n=0}^{\infty}\exp[-(2n+1)(\mu\sigma p)^{\frac{1}{2}}a]. \quad (16)$$

Since this infinite series converges very fast, only the first term will be used. Therefore,

$$h(x',p)=\frac{H_s}{p}\exp[-(\mu\sigma p)^{\frac{1}{2}}x'], \quad (17a)$$

and

$$\frac{dh(x',p)}{dx'}=-H_s\left(\frac{\mu\sigma}{p}\right)^{\frac{1}{2}}\exp[-(\mu\sigma p)^{\frac{1}{2}}x']. \quad (17b)$$

At the boundary $x'=0$,

$$\frac{dh}{dx'}(0,p)=-H_s\left(\frac{\mu\sigma}{p}\right)^{\frac{1}{2}}, \quad (18a)$$

or

$$\frac{dH}{dx'}(0,t)=-H_s\left(\frac{\mu\sigma}{\pi t'}\right)^{\frac{1}{2}}. \quad (18b)$$

If the time instant t' is picked so that $dH/dx'=-(H_0H_s)/x_s$ and the corresponding curve is used to represent the $H-x$ distribution at the instant of time t, the satisfaction of the boundary conditions that H and dH/dx are continuous at $x=x_s$ will be guaranteed. A third condition, namely, $dH/dx=0$ at $x=\delta/2$ or $x'=a$ should be satisfied except for the fact that only one term has been used in the infinite series in Eq. (16).

From Eq. (18b), the curve which gives the correct dH/dx' occurs when

$$t_1'=\frac{\mu\sigma}{\pi}\left(\frac{H_s}{H_0-H_s}x_s\right)^2. \quad (19)$$

To determine the area *abcdfha* (Fig. 5), the first step is to find the area *ahfm* in the following manner:

$$\int_{x_s}^{\delta/2}h(x,p)dx=\int_0^a h(x',p)dx'=\left(\frac{H_s^2}{\mu\sigma p^3}\right)^{\frac{1}{2}}\{1-\exp[-(\mu\sigma p)^{\frac{1}{2}}a]\}\doteq\left(\frac{H_s^2}{\mu\sigma p^3}\right)^{\frac{1}{2}}, \quad (20)$$

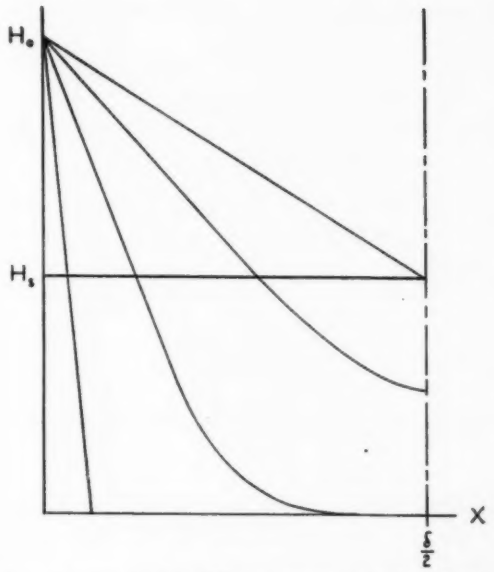


FIG. 4. $H-x$ curves at different instants of time ($\rho=0.5$).

or

$$\int_0^a H(x',t')dx'=\frac{H_s}{(\mu\sigma)^{\frac{1}{2}}}\left\{2\left(\frac{t'}{\pi}\right)^{\frac{1}{2}}\left[1-\exp\left(-\frac{\mu\sigma a^2}{4t'}\right)\right]+(\mu\sigma)^{\frac{1}{2}}a\,erfc\left[\left(\frac{\mu\sigma}{t'}\right)^{\frac{1}{2}}\frac{a}{2}\right]\right\}\doteq 2H_s\left(\frac{t'}{\mu\sigma\pi}\right)^{\frac{1}{2}}. \quad (21)$$

At

$$t_1'=\frac{\mu\sigma}{\pi}\left(\frac{H_s}{H_0-H_s}x_s\right)^2,$$

$$\int_0^a H(x',t')dx'\doteq\frac{2}{\pi}\frac{H_s^2}{H_0-H_s}x_s, \quad (22)$$

which represents the area *ahfm*.

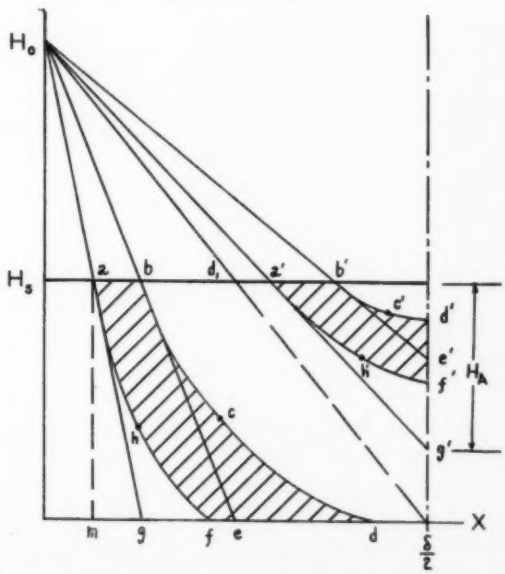


FIG. 5. Area bounded by $H-x$ curves representing flux change per unit height of slab during Δt.

The area *abcdfha*, called A_1, is

$$A_1=H_s dx_s\left[\frac{2-\rho}{2(1-\rho)}+\frac{(4-\pi)\rho}{2\pi(1-\rho)}\right]$$
$$=H_s dx_s[\eta+\kappa] \tag{23}$$
$$=H_s dx_s\left[1+\frac{2}{\pi}\frac{\rho}{1-\rho}\right],$$

where

$$\eta=\frac{2-\rho}{2(1-\rho)},$$
$$\kappa=\frac{(4-\pi)\rho}{2\pi(1-\rho)}.$$

This area, when multiplied by μ, represents the change in flux per unit height of the slab and must equal $E_s dt$. This condition is given by

$$E_s dt=\frac{H_0-H_s}{\sigma x_s}dt=\mu H_s dx_s(\eta+\kappa), \tag{24}$$

or

$$\frac{dt}{\mu\sigma}=\frac{\rho}{1-\rho}(\eta+\kappa)x_s dx_s. \tag{25}$$

By integrating (25), it is found that

$$\frac{t}{\mu\sigma}=\frac{\rho}{2(1-\rho)}(\eta+\kappa)x_s^2=k'x_s^2, \tag{26}$$

and therefore,

$$E_s=\left[\frac{(\eta+\kappa)(1-\rho)B_s H_0}{2\sigma t}\right]^{\frac{1}{2}}. \tag{27}$$

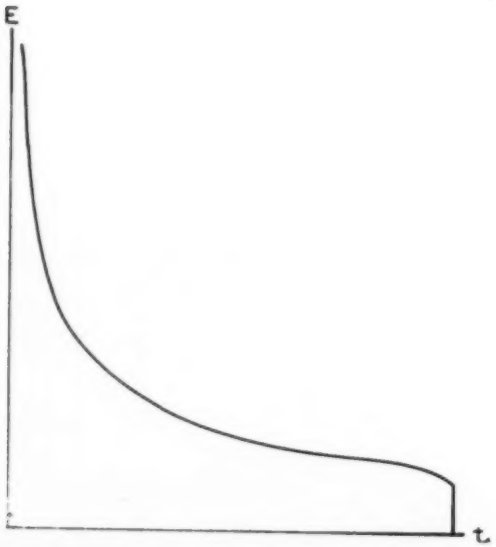


FIG. 6. Wave form of electric field intensity at the boundary ($\rho=0.5$).

The parameter k, defined by Eq. (11), is

$$k=\eta+\kappa=1+\frac{2}{\pi}\frac{\rho}{1-\rho}. \tag{28}$$

Equation (26) holds true until x_s reaches $d_1=(1-\rho)\times\delta/2$ as shown in Fig. 5. The time required for x_s to move from 0 to d_1 is

$$T_1=\frac{\mu\sigma\delta^2}{8}\left[\rho-\frac{\pi-2}{\pi}\rho^2\right]. \tag{29}$$

After x_s reaches d_1, all the preceding arguments hold true with the exception that a straight line with a slope $-(H_0-H_s)/x_s$ will intersect the $x=\delta/2$ line at (H_s-H_A), with H_A being

$$H_A=(H_0-H_s)\left(\frac{\delta/2-x_s}{x_s}\right). \tag{30}$$

Therefore, in deriving an approximate representation of the $H-x$ curves after T_1, the value H_A should be used instead of H_s in Eq. (17). The $H-x$ curves after T_1 are given by

$$h(x',p)=\frac{(H_s-H_A)}{p}+\frac{H_A}{p}\exp[-(\mu\sigma p)^{\frac{1}{2}}x']. \tag{31}$$

Following the same procedure indicated by Eqs. (17), (18), the curve which gives the correct $\partial H/\partial x'$ occurs when

$$t_2'=\frac{\mu\sigma}{\pi}\left(\frac{H_A}{H_0-H_s}x_s\right)^2. \tag{32}$$

The area $a'b'c'd'f'h'a'$, called A_2, is

$$A_2=H_A\left[1+\frac{2}{\pi}\frac{H_A}{H_0-H_s}\right]dx_s. \tag{33}$$

The area A_2, when multiplied by μ, must equal $E_s dt$:

$$\mu H_A\left[1+\frac{2}{\pi}\frac{H_A}{H_0-H_s}\right]dx_s=\frac{H_0-H_s}{\sigma x_s}dt. \tag{34}$$

By substituting (30) into (34), the result gives

$$\frac{dt}{\sigma\mu}=\frac{\delta^2}{2\pi}\frac{dx_s}{x_s}-\frac{\delta(4-\pi)}{2\pi}dx_s-\frac{\pi-2}{\pi}x_s dx_s. \tag{35}$$

The time interval required for x_s to move from d_1 to $\delta/2$ may be found by integrating (35) and substituting d_1 and $\delta/2$ as the lower and upper limits of integration. The time interval T_2 is

$$T_2=\mu\sigma\delta^2\left[\frac{1}{2\pi}\log\frac{1}{1-\rho}-\frac{1}{2\pi}\rho+\frac{\pi-2}{8\pi}\rho^2\right]. \tag{36}$$

The total "switching time" is thus found to be

$$T_s = T_1 + T_2 = \frac{\mu\sigma\delta^2}{8}\left[\frac{4}{\pi}\log\frac{1}{1-\rho} - \left(\frac{4}{\pi}-1\right)\rho\right]. \quad (37)$$

The switching time T_s is plotted against ρ in Fig. 2. It checks with the results of the painstaking graphical method. The deviation of (37) from the graphical result, in the author's opinion, is mainly caused by the approximations used in Eqs. (16), (20), and (21) in order to give readily recognizable results. The approximation to Eq. (16) tends to give a higher value of t_1'. This, together with the simplification used in the integrations in Eqs. (20) and (21), tends to give a higher value of change in flux and therefore slows down the moving of x_s. Therefore, the switching time given by Eq. (37) is higher than the actual value.

The wave form of the electric field intensity, calculated from Eq. (27) and indirectly from (35), is shown in Fig. 6 for the value of ρ equal to 0.5.

CONCLUSION

In the paper the rigorous solution of the nonlinear diffusion equation for the case $\delta = \infty$ (see Appendix) is adapted to the practical case δ = finite. The switching time T_s is expressed by a simple formula, Eq. (37). The wave form of the electric field intensity E_s is shown in Fig. 6.

ACKNOWLEDGMENT

The author is indebted to Professor E. Isaacson of New York University for his helpful suggestions.

APPENDIX

Rigorous Solution When δ Approaches Infinity

When the semi-infinite space is divided into saturated and unsaturated regions:

$$H_I = H_0 - \alpha x, \quad (38)$$

$$H_{II} = A\,erfc\left[\left(\frac{\mu\sigma}{t}\right)^{\frac{1}{2}}\frac{x}{2}\right]. \quad (39)$$

where **I** denotes the saturated region and **II** denotes the unsaturated region. Equation (39) is the solution if and only if A is a constant.

At the boundary x_s, H is continuous:

$$H_I = H_0 - \alpha x_s = H_s, \quad (40)$$

$$H_{II} = A\,erfc\left[\left(\frac{\mu\sigma}{t}\right)^{\frac{1}{2}}\frac{x_s}{2}\right] = H_s. \quad (41)$$

Also, $\partial H/\partial x$ is continuous:

$$\partial H_I/\partial x = -\alpha, \quad (42)$$

$$\partial H_{II}/\partial x = -A\left(\frac{\mu\sigma}{\pi t}\right)^{\frac{1}{2}}\exp(-\mu\sigma x_s^2/4t) = -\alpha. \quad (43)$$

From Eq. (41), it can be concluded that

$$t/\mu\sigma = k' x_s^2, \quad (44)$$

and Eq. (41) becomes

$$A\,erfc\frac{1}{2(k')^{\frac{1}{2}}} = H_s. \quad (45)$$

On combining (40), (43), (44), and (45), it is found that

$$(k')^{\frac{1}{2}}\exp(1/4k')\,erfc\left[\frac{1}{2(k')^{\frac{1}{2}}}\right] = \frac{H_s}{(\pi)^{\frac{1}{2}}(H_0 - H_s)} = \frac{\rho}{(\pi)^{\frac{1}{2}}(1-\rho)}. \quad (46)$$

The parameter k', which is the solution of Eq. (46), gives the complete solution of the problem. Equation (46) shows that k' is a constant and therefore A is a constant by Eq. (45).

The asymptotic expansion of (46) shows that when k' approaches infinity,

$$k' = \frac{1}{\pi}\left(\frac{\rho}{1-\rho}\right)^2; \quad (47)$$

and when k' approaches zero,

$$k' = \frac{\rho}{2(1-\rho)}. \quad (48)$$

In the approximate method, the corresponding value of k', given by Eq. (26), is

$$k' = \frac{\rho}{2(1-\rho)}\left[1 + \frac{2}{\pi}\frac{\rho}{1-\rho}\right]. \quad (49)$$

When ρ approaches unity,

$$k' = \frac{1}{\pi}\left(\frac{\rho}{1-\rho}\right)^2; \quad (50)$$

and when ρ approaches zero,

$$k' = \frac{\rho}{2(1-\rho)}. \quad (51)$$

Equations (50) and (51) are the same as Eqs. (47) and (48), respectively. For values of ρ between 0 and unity, the values of k' given by Eqs. (46) and (49) differ by less than 6 percent.

Cathode Effects in the Dielectric Breakdown of Liquids

J. K. Bragg, A. H. Sharbaugh, and R. W. Crowe
General Electric Research Laboratory, Schenectady, New York

(Received August 14, 1953)

The apparent (measured) electric strength of a liquid dielectric depends on the nature of the cathode used. The field emission of electrons from the metallic cathode is assumed to be responsible for this effect; it may provide space charge which distorts the electric field in the dielectric. The nature of the cathode determines the range of electric fields over which the emission becomes appreciable.

In this paper we discuss in detail the emission of charge and the formation of space charge. Because of theoretical and experimental uncertainties connected with supposedly uniform metallic surfaces, an electrode consisting of an electrolyte solution is also considered. Although the experimental results demonstrate that *ions* are emitted from such cathodes under the influence of strong electric fields, the role in providing space charge is similar to that of metal electrodes which emit electrons.

For the former electrode, the considerations regarding space-charge formation predict a definite dependence of apparent electric strength on electrolyte concentration and provide a way of testing the basis of the theory. Certain experiments reported here confirm the predictions of the theory.

I. INTRODUCTION

THE influence of the nature of the cathode metal upon the breakdown of dielectrics has been observed by many investigators. von Hippel and Alger[1] first attributed this effect, as observed in the study of alkali-halide crystals, to field emission of electrons from the cathode. Evidence that the origin of the effect in liquids is also field emission was found by Salvage,[2] who succeeded in showing a rough correlation between his experimental results and the vacuum work function of the metal used as cathode. Because of the sensitivity of the electron-emission characteristics of a metal to the physical and chemical condition of its surface, however, such correlations are at best qualitative, and may be difficult to reproduce.

We have tried to put the suggestion of von Hippel in a quantitative form, and have designed a new kind of breakdown experiment to give a reliable test of the resulting theory. The experiment involves the use of a cathode consisting of an aqueous solution of electrolyte; the emission of ions from this cathode is a reproducible phenomenon whose dependence upon electrolyte concentration can be predicted. The theory depends on a distortion of the electric field which may be caused by the emission of charge from the cathode into the dielectric.

For example, if an electric field of 10^6 v/cm is impressed on a dielectric, field emission of electrons from a metal cathode may occur as shown by LePage and DuBridge.[3] This emission may provide a considerable density of negative charge in the gap between electrodes. The amount of such charge which can be transported across the gap to the anode is limited by the properties of the dielectric (the mobility it allows the charges, and the dielectric constant) and by the geometry of the gap. The current that flows is, therefore, dependent on cathode properties at low fields, but when the field is high enough to cause strong emission, the current becomes independent of the nature of the cathode, and is prescribed by the properties of the gap through which the charge must be transported.

In the latter case, a nonuniform distribution of space charge arises, and the electric field in the gap becomes inhomogeneous. It is to this feature that the cathode effects are ascribed; that is, under these "space-charge-limited" conditions, the electric field in a region of the dielectric is enhanced to a magnitude considerably greater than the average (measured) field. The value of the average field at which this occurs is, however, *dependent on the nature* (*emission characteristics*) of the cathode.

The connection of this with breakdown strength measurements must be made with the help of a somewhat arbitrary assumption. We first suppose that the *intrinsic electric strength* of a dielectric is a meaningful property. We define it in terms of the magnitude of the homogeneous electric field, existing in a region of the dielectric, necessary to disrupt this region or establish a conducting path across it, initiating processes from outside the region being excluded. The region considered must be large enough so that the property so defined is independent of the size of the region.

The intrinsic electric strength, defined in this way, obviously depends only on the nature of the dielectric; what is usually measured does not. Its relation to cathode effects is the subject of our basic assumption. We assume that a dielectric will break down whenever the average electric field over a region of dimension δ in the direction of the field becomes equal to the intrinsic electric strength, the magnitude of δ being just that for which the homogeneous electric field necessary for disruption of the region becomes independent of δ.

It should be noticed that this assumption is independent of the nature of the breakdown process, and all the conclusions we draw from it will apply whatever the breakdown mechanism. However, the understand-

[1] A. von Hippel and R. S. Alger, Phys. Rev. **76**, 127 (1949).
[2] B. Salvage, Proc. Inst. Elec. Engrs. **98**, 227 (1951).
[3] W. R. LePage and L. A. DuBridge, Phys. Rev. **58**, 61 (1940).

ing of breakdown mechanisms and the techniques of measurement are not yet sufficiently advanced to make this hypothesis operationally useful. We will consequently use a somewhat indefinite idea; we ask simply that the electric field become equal to the intrinsic strength somewhere in the dielectric.

In the remainder of this paper, we will develop the consequences of the assumptions set forth above and report the extent to which these ideas are supported by experiment.

II. MECHANISMS OF CHARGE EMISSION

A. Field Emission

Evidence that the high field-emission current from metal cathodes into dielectrics is essentially field emission has been presented by LePage and DuBridge.[3] They measured the emission current into toluene as a function of field strength and temperature. Figure 1 shows some of their results, plots of the logarithm of the current against reciprocal temperature for various field strengths. It will be observed that the magnitudes of the slopes decrease as the field strength is increased. This suggests the gradual transition from Schottky emission, which has a pronounced exponential temperature dependence, to field emission, which is independent of temperature. The electric fields with which we deal are greater than any listed in this diagram.

Field emission into the vacuum is described fairly well by the Fowler-Nordheim equation,[4]

$$j = f(\chi)E^2 \exp(-b\chi^{\frac{3}{2}}/E). \quad (1)$$

Here b is a constant, χ is the vacuum work function, and E is the electric field. The quantity $b\chi^{\frac{3}{2}}$ is of the order of 10^8 if E is expressed in volts per centimeter, so that this expression describes a current which increases very strongly in the region around 10^7 v/cm. $f(\chi)$ is usually of such a magnitude that appreciable emission appears at about 10^7 v/cm. The applicability of the law depends on the presence of uniform surfaces; in general, the presence of points and patches must be taken into account.

If the emission is into a dielectric, the situation is not so clear. If the dielectric may be regarded as structureless, one may derive the equation[3]

$$j = f^*(\chi^*)E^2 \exp(-b^*\chi^{*\frac{3}{2}}/E). \quad (2)$$

The constants f^* and b^* now involve the dielectric constant ϵ of the dielectric, and χ^* is the work function for the cathode-dielectric system, i.e., the difference between the Fermi level of electrons in the cathode and the electrostatic potential of an electron removed to infinity in the dielectric. It can be shown that $\chi^* < \chi/\epsilon$, but beyond this, χ^* can be specified only by measurement.

[4] R. H. Fowler and L. Nordheim, Proc. Roy. Soc. (London) A119, 173 (1928).

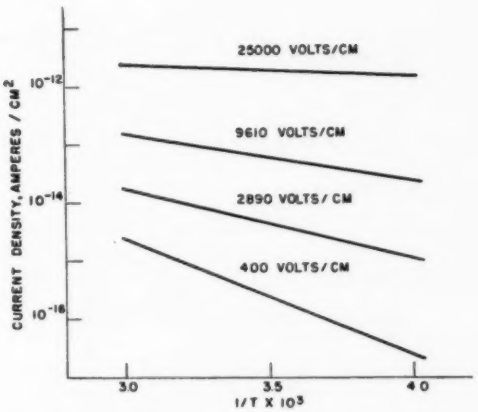


FIG. 1. Currents in toluene at various field strengths as function of temperature (after LePage and DuBridge).

In actual practice, the equation just given probably does not apply at all. If the dielectric is crystalline so that its electron states are distributed in bands, then a different kind of barrier to emission arises, whose shape depends on the concentration and nature of impurities in the crystal.[5] Something similar may occur in amorphous solids and liquids; furthermore, the mean free path may be so short in these substances that the whole basis for the derivation of the electron emission laws is invalidated.

Nevertheless, the field-emission current should depend on the field in an exponential way, and the equation

$$j = a \exp(-b/E),$$

with empirical constants, should be satisfactory.

B. Ion Emission

Experimental evidence to be described in this paper indicates that, under proper conditions, an emission current of ions may be obtained. The experiments involve the use of aqueous electrolyte solutions as cathodes in the study of the breakdown of liquids immiscible with water. We wish to discuss here the nature of the emission, dealing mainly with a simple picture of the process which has to be modified to agree with experiment, but which must be understood in order to interpret the actual results.

In Fig. 2 we have sketched the "motive" (electrostatic potential plus an effective potential caused by image forces) of a negative charge in a system consisting of aqueous electrolyte cathode, a dielectric liquid, and a metal anode, in the presence of an electric field. The slight decrease in the motive just to the left of the electrolyte-dielectric interface is due to the distortion of the ionic distribution by the electric field; in contrast to what occurs in metallic conductors, the field must extend an appreciable distance into the cathode. The

[5] See, for example, H. C. Torrey and C. A. Whitmer, *Crystal Rectifiers* (McGraw-Hill Book Company, Inc., New York, 1948).

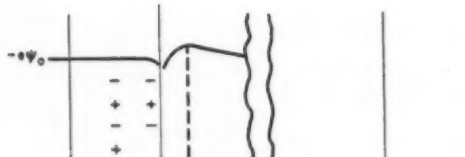


FIG. 2. Motive of negative charge in electrolyte dielectric system.

rise to the right of the interface is due to an assumed increase in energy of the ions as they move from the water layer to the dielectric layer, and the shape of the curve in this region is prescribed by the image force between the ion and the electrolyte.

Calculation shows that for as low a concentration of ions as 10^{-4}N, the drop in potential in the electrolyte is less than 0.25 v out of 10 kv impressed across 0.01 cm of dielectric. This drop will be ignored henceforth.

The difference in energy of a univalent ion in water and in benzene may be of the order of 2 ev. To illustrate, we adopt the very simple view that the energy of an ion in a dielectric liquid is just that required to charge the ion in the medium[6]

$$E_\epsilon = -e^2/2\epsilon r_0, \tag{3}$$

(e=ionic charge, ϵ=dielectric constant, r_0=ionic radius); then the energy difference is

$$e\chi_0 = \frac{e^2}{2r_0}\left(\frac{1}{\epsilon_1}-\frac{1}{\epsilon_2}\right) \approx \frac{e^2}{2\epsilon_1 r_0}; \quad \epsilon_2 \gg \epsilon_1. \tag{4}$$

Although this is obviously very much oversimplified, we may obtain a certain degree of consistency by using ionic radii as determined from heats of solution with the aid of this formula.[6] This procedure is still not quantitative, because we have to use, for the ion in the dielectric, the radius determined from aqueous solutions.

For a typical ion, Cl^- (r_0=2.13A), we find $e\chi_0$=1.4 ev for the water-benzene transfer.

There is the possibility that the ion may enter the dielectric still associated with several water molecules. If the resulting droplet has a radius r, the energy difference for the transfer is

$$e\chi = \frac{e^2}{2\epsilon_1 r} + 4\pi r^2\sigma, \tag{5}$$

where σ is the water-benzene interfacial free energy per unit area. The value of r which makes this expression a minimum is

$$r^\ddagger = (e^2/16\pi\epsilon\sigma)^{\frac{1}{2}} \approx 3.8\text{A},$$

a number about double the ionic radius. Despite the uncertainty in using an interfacial free energy for a droplet so small, this makes it probable that a few molecules of water are carried along with the ion. Experimental evidence that this is so will be presented later. The energy difference thus obtained is about $e\chi^\ddagger$=1.2 ev.

The quantity $e\chi_0$ (or $e\chi^\ddagger$) plays the role of a work function for the ion. The motive for the ion is of the same form as that for an electron in a dielectric opposite a metal cathode, as evaluated by LePage and DuBridge.[3] The height of the maximum in the motive above the value in the electrolyte is

$$e\chi_m = e\chi^\ddagger - \frac{e^{\frac{1}{2}}E^{\frac{1}{2}}}{\epsilon^{\frac{1}{2}}}. \tag{6}$$

The concentration of ions at x_m (see Fig. 2), provided they are supplied sufficiently rapidly from the cathode, is

$$n(x_m) = n_0 \exp(-e\chi^\ddagger/kT) \exp\left(\frac{e^{\frac{1}{2}}E^{\frac{1}{2}}}{\epsilon^{\frac{1}{2}}kT}\right); \tag{7}$$

here n_0 is the concentration of ions in the electrolyte. This is similar to the Schottky correction for the effect of a finite collecting field on the barrier to thermionic emission into a vacuum.

A difference arises when we try to calculate the emission current. In the present case the ions cross the barrier by diffusion. There is no way of taking this into account precisely, so an approximate assumption which is exceedingly simple will be made. We suppose the ions move across the barrier with the drift velocity μE they would have in the presence of the field E alone. Then we have for the current density

$$j = -\mu e E n_0 \exp(-e\chi^\ddagger/kT) \exp\left(\frac{e^{\frac{1}{2}}E^{\frac{1}{2}}}{\epsilon^{\frac{1}{2}}kT}\right). \tag{8}$$

This approximation is exact if applied to the simplified problem of the barrier shown in Fig. 3, which can be solved for the motion of ions in the combined electric and concentration gradients. The result is that the concentration at the top of the barrier, multiplied by the drift velocity of the charges in the electric field alone, gives the current across the barrier.

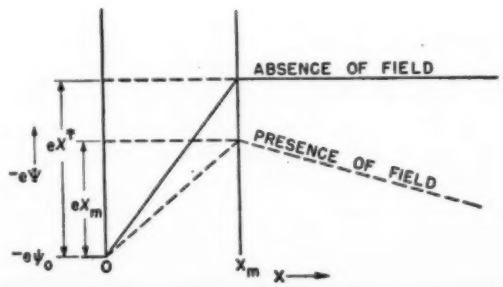


FIG. 3. Potential barrier for simplified emission problem.

[6] M. Born, Z. Physik I, 45 (1920).

With sufficient labor one can obtain arbitrarily accurate numerical solutions for a barrier of any complexity, but the qualitative conclusion we want to establish here does not demand such refinements, nor does our physical knowledge of the situation justify them.

Equation (8) predicts a negligible emission current at any reasonable field strength. The reason is that the Schottky correction, $e^{\frac{1}{2}}E^{\frac{1}{2}}/\epsilon^{\frac{1}{2}}kT$, lowers the barrier by only 0.25 v for a field of 10^6 v/cm. This implies that the barrier layer is about 25A thick; a barrier layer of 100A, on the other hand, gives a correction to $e\chi^{\frac{1}{2}}$ of 1 v and an emission current of the order of 0.5 amp.

We have obtained evidence of emission currents of almost this order of magnitude, but only after one or more preparatory electric pulses. Apparently the function of the preparatory pulses is to blur the Schottky layer, thus providing enhanced emission. We will simply make the obvious assumption that such an emission current depends exponentially on the electric field.

III. SPACE-CHARGE EQUATIONS

A. Simple Space-Charge Equations

The charge and field distributions in the gap (assumed to consist of two infinite plane conductors separated by a distance a centimeters) are governed by the following relationships:

$$\begin{aligned} &\epsilon(d^2\varphi/dx^2)=-4\pi\rho; \\ &j=-\rho\mu(d\varphi/dx), \quad 0\leqslant x\leqslant a. \end{aligned} \tag{9}$$

In these equations ϵ is the dielectric constant, φ the electrostatic potential, μ the charge mobility, j the current density, ρ the charge density, and x the coordinate (which varies from 0 to a across the gap). We consider only negative charge, so that μ and ρ are negative. For the moment we assume that the mobility is independent of the electric field. The consequences of any failure of this assumption will be assessed shortly.

The problem is made definite by the statement of certain boundary conditions:

$$\begin{aligned} \varphi&=0 \quad \text{at} \quad x=0, \\ \varphi&=\varphi_0 \quad \text{at} \quad x=a, \\ -(d\varphi/dx)&=E_c \quad \text{at} \quad x=0. \end{aligned} \tag{10}$$

The first two conditions always obtain. The simplest assumption about the third is that $E_c=0$, and the solutions corresponding to this boundary condition are the following:[7]

$$\begin{aligned} &\varphi=\varphi_0 x^{\frac{1}{2}}/a^{\frac{1}{2}}, \\ &\rho=(3/32\pi)(\epsilon\varphi_0/a^{\frac{1}{2}}x^{\frac{1}{2}}), \\ &j=(9\epsilon\mu\varphi_0^2/32\pi a^3). \end{aligned} \tag{11}$$

[7] See, for example, J. D. Cobine, *Gaseous Conductors* (McGraw-Hill Book Company, Inc., New York, 1941), p. 128.

TABLE I.

γ_c	0	0.20	0.40	0.60	0.80	1.00
η	2.25	2.14	1.84	1.35	0.70	0.00
γ_a	1.50	1.48	1.41	1.31	1.16	1.00

These are analogs of the Langmuir-Childs equations, valid, however, for liquids, solids, and high-pressure gaseous dielectrics. A consequence of the first of these is that the electric field at the anode, E_a, is the average field $E_0=\varphi_0/a$ increased by a factor 3/2.

B. The Self-Consistent Solution

One of the results of the simple model is that a finite current flows through the gap. The solution is thus obviously not self-consistent, since the assumption $E_c=0$ implies that the field-induced emission current is zero. The boundary condition must be so adjusted that the current flowing in the gap, as derived from the space-charge equations, is equal to the emission current of the cathode given by its field-emission law.

We must, therefore, solve Eqs. (9) under the general boundary condition $-(d\varphi/dx)_{x=0}=E_c$. We will then regard E_c as a function of j in the solutions. The integrations, together with appropriate determination of the constants arising, give now

$$\varphi=\frac{2}{3}\frac{\epsilon\mu}{8\pi j}\left[\left(\frac{8\pi j}{\epsilon\mu}x+E_c^2\right)^{\frac{3}{2}}-E_c^3\right]. \tag{12}$$

This and succeeding results reduce to the simple space-charge equations if we set $E_c=0$.

The current is derived by putting φ equal to φ_0 when x equals a. The implicit equation for j is then

$$\varphi_0=\frac{2}{3}\frac{\epsilon\mu}{8\pi j}\left[\left(\frac{8\pi aj}{\epsilon\mu}+E_c^2\right)^{\frac{3}{2}}-E_c^3\right]. \tag{13}$$

In terms of convenient dimensionless quantities $\gamma_c=E_c/E_0$ and $\eta=8\pi aj/\epsilon\mu E_0^2$, this equation becomes

$$\eta=\tfrac{2}{3}[(\eta+\gamma_c^2)^{\frac{3}{2}}-\gamma_c^3]. \tag{14}$$

Table I gives solutions of this cubic equation for η, for various values of γ_c. In addition to the reduced current η, the table lists values of the reduced anode field $\gamma_a=E_a/E_0$ obtained by differentiating Eq. (12) with respect to x and then setting x equal to a. As previously, if the cathode field is made to be zero, the anode field is three-halves of the average field. On the other hand, if γ_c is 1.0, then γ_a is 1.0 also.

Equation (14) gives a relation between the reduced current and the reduced cathode field which arises from consideration of the transport of charge through the dielectric. There is another equation connecting them, namely the cathode emission law

$$\eta=\eta(\gamma_c, E_0). \tag{15}$$

For a given average field E_0, Eqs. (14) and (15) can be solved simultaneously to give numerical values of the reduced current and cathode field, and hence the reduced anode field.

In Fig. 4 we have drawn a typical cathode-emission characteristic, and the space-charge-limited current as a function of field strength, to illustrate the meaning of these results. The assumed cathode characteristic has the form given by Eq. (15) and is plotted against E_0. In the region $E_0<E_1$, emission is so feeble that $\gamma_c\approx 1$, and the field is essentially undistorted. Note that our self-consistent solution follows the cathode characteristic in this region. At some field E_1 the reduced emission current has grown to, say, 0.2, and Table I shows that appreciable deviations of γ_a and γ_c occur. As E_0 is further increased, the strong tendency of the emission current to increase rapidly forces the system over into space-charge-limited conditions under which the reduced current in the gap is independent of the field (the actual current varies as the square of E_0).

Comparison of these results with those of the preceding section shows that the simple treatment of the problem gives values of the anode field which are asymptotically correct at low and high fields. Consequently, the discussion of the breakdown process to be given later will be developed on the basis of the simple space-charge laws.

C. Dependence of the Mobility on the Electric Field

In the discussion just given we have assumed that the mobility of the negative charges is independent of the electric field. This assumption is undoubtedly defective in the field strength region treated here. Nevertheless, there is no information about the field dependence of electron drift velocities in liquids under high fields. Shockley[8] has used the work of Seitz[9] on the mobility of electrons in crystalline, nonpolar insulators to show that in this case the mobility ought to decrease with electric field at high fields, eventually varying as $E^{-\frac{1}{2}}$. On the other hand, it is possible that an electron in a nonpolar *liquid* may become trapped or "solvated," and thus behave like a molecular ion, in which case the mobility would be expected to increase with field at high fields (see below). It can be proved that if the mobility varies as $E^{-\frac{1}{2}}$, the anode field under space-charge-limited conditions is 5/3 times the average field.

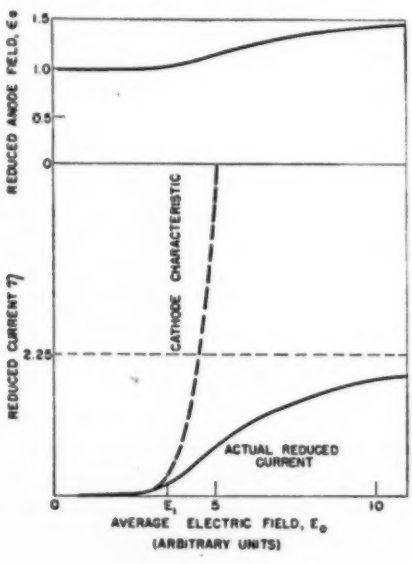


FIG. 4.

For ionic charge carriers the matter is somewhat clearer. It is certain that the drift velocity of such charges will increase strongly with field in the high field region. Not so certain is the functional form, but a crude approximation to the current, often used, is

$$j=a\rho\sinh(\beta d\varphi/dx);\quad a,\beta>0. \tag{16}$$

This current deviates from an ohmic law if $\beta(d\varphi/dx)$ is not small compared to unity. If φ_0/a (the average field) is *large* compared to $1/\beta$, it can be shown that the field in the gap is essentially undistorted by space charge.

Because of uncertainties in both emission and conduction laws, it is not worth while to develop intermediate cases. We may conclude, however, that in practice the extent of distortion of the electric field in the gap due to ionic space charge is expressed by a factor lying between 1 and 1.5, multiplying the average field. On the other hand, an electronic space charge may produce a distortion of field expressed by a factor lying between 1 and 2.

IV. DEPENDENCE OF MEASURED ELECTRIC STRENGTH ON CATHODE CHARACTERISTICS

In the previous sections, we have described possible kinds of field-dependent charge emission, and the space-charge effects which they may produce. We have still to show the connection between these and measurements of electric strength. In the detailed discussion to follow, we will use principally the results of the simple space-charge calculation. The modification that may be made by a more refined treatment will be indicated afterward.

In Fig. 5 we have plotted some current densities *versus* electric field strength. The solid lines represent the emission characteristics of a hypothetical series of cathodes as a function of the field strength at the cathode surfaces; the emission described by these curves is supposed to increase exponentially with electric field. The dotted curve is the space-charge-limited current plotted against the average field in the gap.

Consider first the cathode curve marked (I). If such a cathode is used, an increase in the applied electric field will bring about an increase in the emission current, following along the solid curve until it intersects the

[8] W. Shockley, Phys. Rev. 82, 330 (1951).
[9] F. Seitz, Phys. Rev. 76, 1376 (1949).

dotted one. Beyond this point, the potential emission current of the cathode is larger than can be collected. The actual current, therefore, follows the dotted line thenceforth; the emission is "space-charge-limited," and the field in the gap is no longer uniform. The simple space-charge law states that, in this situation, the maximum field in the gap (that at the anode) is three-halves the average field. Thus, when we have progressed to the right in Fig. 5 until the average field is two-thirds the intrinsic strength, breakdown will occur. This is true of any cathode whose characteristic intersects the dotted curve at a field less than $(2/3)E_i$, where E_i is the intrinsic strength of the dielectric.

Next we consider the cathode curve marked (III). The emission current from this cathode, at the field E_i, is still below the space-charge-limited current. Consequently, the field on the cathode and throughout the gap is E_i, and breakdown occurs at the average field E_i.

Finally, when the curve (II) intersects the dotted line and the field becomes distorted, a field greater than E_i is produced at the anode, so that breakdown occurs when the average field corresponds to the point of intersection. We can describe the entire situation in

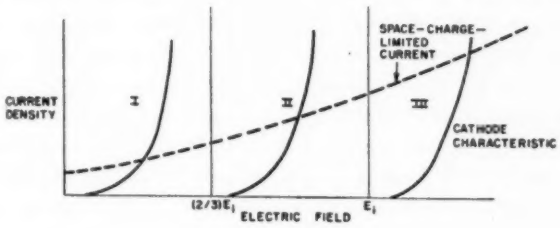


FIG. 5.

the following approximate way: any cathode which begins to emit strongly at fields below $(2/3)E_i$ gives breakdown at $(2/3)E_i$. A cathode which begins to emit at a field between $(2/3)E_i$ and E_i gives breakdown at that field, while a cathode whose emission is negligible below E_i gives breakdown at E_i. The resulting plot of breakdown strength *versus* cathode nature (effective work function of a metal increasing to the left, or the concentration of an aqueous electrolyte cathode increasing to the right) is shown in Fig. 6, and is characterized by two cathode-independent regions (the upper and lower flat portions) and a transition region.

The equation of the transition curve is obtained by equating emission current and space-charge-limited current; if we treat the latter as constant compared to the rapidly changing emission current, we obtain, in the case of ion emission,

$$n_0 \exp(aE^\theta) = \text{constant},$$

or

$$E^\theta = 1/a(\text{constant} - \log n_0). \qquad (17)$$

Here θ is a constant near unity (if the emission were of the Schottky type, it would be one-half). Thus, the breakdown field should vary approximately linearly

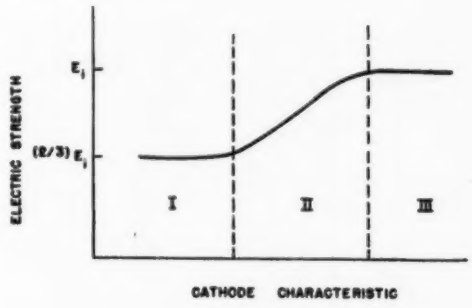


FIG. 6.

with the logarithm of the concentration in the intermediate portion of the graph of Fig. 6.

One modification of the results of this section, introduced by more detailed space-charge considerations, should be mentioned. The relation which the upper and lower flat portions of the curve of Fig. 6 bear to each other depends on the law relating the drift velocity of the charge carrier to the electric field. In Sec. III, we saw that for electrons this relationship is not understood, so that no predictions can be made about the ratio of the two flat portions if the cathodes used are metals.

On the other hand, if the cathode emits ions, we can say with confidence that the drift velocity varies at least as strongly as the first power of the field. So, according to the results of Sec. III, the ratio of the fields given by the upper and lower flat portions must lie between one and three-halves.

V. EXPERIMENTAL METHODS

In practice, we are usually concerned with metallic cathodes, but any detailed theory of their effects is difficult to develop and impossible to verify experimentally at present. We have developed a theory for the ion-emitting cathode because it runs parallel to that for an electron emitter, and at the same time, is susceptible to a detailed experimental study. We will now report the results of our experiments with such ion-emitting cathodes.

A. Experimental Techniques

The experiments were carried out with a rectangular pulse generator which delivers single, well-defined, rectangular voltage pulses of variable amplitude and duration. Amplitudes up to 15 kv may be obtained, with durations ranging from 0.25 to 54 μsec. The use of such pulses should essentially eliminate erroneous results due to local, prebreakdown heating of the spark gap. This effect, giving rise to what is known as "thermal breakdown," has often been observed in systems to which dc and ac voltages were applied. This is obviously undesirable if one is attempting to study the fundamental mechanism of breakdown.

The construction details of the breakdown cell are shown in Fig. 7. The general design is similar to that

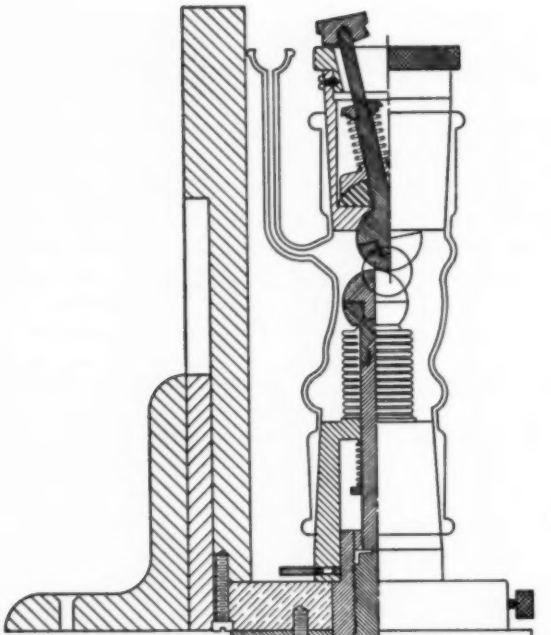


FIG. 7. Assembly of breakdown cell.

used by Bahre,[10] except that one electrode is made movable through the use of a bellows and micrometer drive arrangement. The cell is equipped with optical windows to make it possible to monitor the breakdown during an experiment, and to adjust the gap spacing by optical means. The ball and socket arrangement for the upper electrode is provided so that a single solid electrode may be used for a number of experiments, a fresh part of its surface being used for each one. For the present series of experiments, the cell was used in the vertical position, and a glass cup filled with aqueous electrolyte substituted for the lower electrode.

Benzene was chosen as the dielectric for the measurements because of its ready availability in a relatively pure condition, and because of its immiscibility with water. The material used was Baker and Adamson (cp) thiophene-free benzene; no attempt was made to purify it further. The benzene was filtered through an ultrafine (0.9–1.4μ pore size) fritted glass filter to remove solid particles.

After assembly of the dielectric-electrode system, the travelling microscope (32×) was focused on the gap through the optical windows of the cell (Fig. 8). The gap was illuminated with the microscope illuminator. The electrode spacing of 0.0084±0.0001 cm was set using the interference colors which appeared when the electrode separation was very small; we used a particular red color which corresponded to a separation of 0.0008±0.0001 cm. From this point, the spacing was increased by 0.0076 cm, as read from the micrometer, to its final value of 0.0084 cm.

Two techniques were used for making the electric strength measurements. The first, hereafter referred to as the "many-pulse" technique, consisted of the application of single, constant-length, rectangular pulses of gradually increasing amplitude, until disruption of the dielectric occurred. In the second method, only one pulse, near breakdown, was applied to the cathode-dielectric-anode system. If breakdown did not occur, a pulse of slightly higher voltage was applied to the next, completely new system. If breakdown resulted, a pulse of slightly lower voltage was applied to the next system. In the present investigation, this method served to demonstrate an effect to be discussed in later sections.

B. Correction for Deformation of the Electrode Surface

When a voltage is applied to a breakdown gap having a liquid electrode, it causes a deformation of the electrode surface whose extent depends on magnitude and duration of the voltage. A method of correction for this deformation has been described;[11] it involves the use of voltage pulses of several different lengths. For systems involving the mercury cathode, it was found that the decrease in electrode separation resulting from such deformation is linear with pulse length over a considerable range of pulse lengths in excess of one microsecond, giving rise to a linear dependence of electric strength on pulse length in this region. A linear extrapolation of the electric strength to zero time gives a value approximately corrected for electrode deformation.

The apparent electric strength of benzene was measured as a function of pulse length, using aqueous lithium chloride solutions as cathodes; the results are given in Fig. 9. The curves are not far from linear for pulse lengths from 0.7 to 4.7 microseconds, but their slopes indicate a more pronounced deformation than was realized with the mercury cathode. (The rapid rise below 0.7 μsec is a property of the dielectric itself, not of importance here.) Since the nature of the deformation in the pulse-length region below one microsecond is not known, a linear extrapolation of the curve to zero time may yield a result which is somewhat in error, especially because electrode distortion is large. It should also be mentioned that the electric strengths are those of benzene saturated with water, as will always be the case when an aqueous solution is used as an electrode.

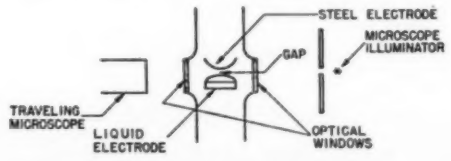


FIG. 8. Assembly for setting gap.

[10] W. Bahre, Arch. Elektrotech. **31**, 141 (1937).

[11] Sharbaugh, Crowe, and Bragg, J. Appl. Phys. **24**, 814 (1953).

VI. EXPERIMENTAL RESULTS AND DISCUSSION

A. The Many-Pulse Technique

The principal purpose of the present study is the verification of the theory of cathode effects which we have described. According to the theory, a plot of the electric strength *versus* the logarithm of the ion concentration should consist of three parts: (1) a level portion at high concentrations, (2) a linear increase over an intermediate concentration range, and (3) another level portion at low concentrations.

This effect has been demonstrated using lithium chloride solutions as cathodes (the high solubility of lithium chloride in water makes it possible to cover a wide range of concentrations). The measurements were made with voltage pulses of length 1.8 microseconds, and corrected to zero time as described in the preceding section. The data in Fig. 9 show that the value obtained in this manner exceeds by about 15 percent that measured with a pulse length of 1.8 microseconds.

The electric strengths measured in this study are plotted in Fig. 10 against the logarithm of the ionic concentration of the cathode. The hollow circles give the data obtained by the many-pulse technique. The two level parts of the curve, together with the connecting linear portion, are clearly evident from these data. The significance of the single-pulse results, represented by the filled circles in Fig. 10, will be discussed in a later section.

If our interpretation of the results in Fig. 10 is correct, the value of the field strength corresponding to the upper flat portion of the curve has a special significance. It is the electric strength measured in the absence of appreciable space charges, and corresponds to the *intrinsic* strength of the dielectric. Unfortunately, for reasons which we have already given, the result has no absolute significance; this will be the case until an extrapolation procedure free from arbitrary assumptions is developed. Nevertheless, the relative values reported here are quite precise, and serve to support the basic ideas involved in the theory of cathode effects.

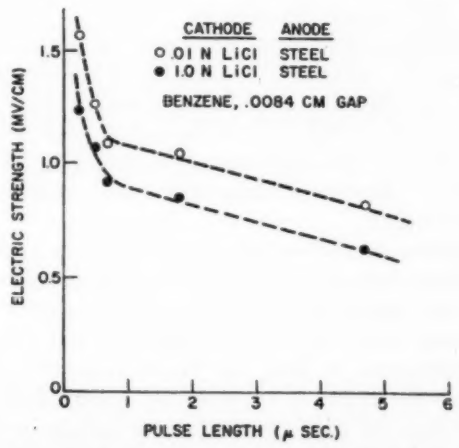


FIG. 9. Dependence of the electric strength of benzene upon pulse length.

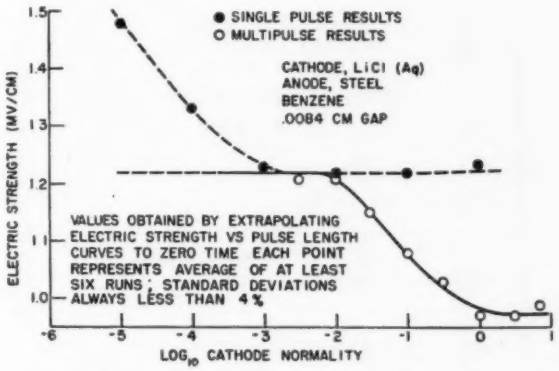

● SINGLE PULSE RESULTS
○ MULTIPULSE RESULTS
CATHODE, LiCl (Aq)
ANODE, STEEL
BENZENE
.0084 CM GAP
VALUES OBTAINED BY EXTRAPOLATING ELECTRIC STRENGTH VS PULSE LENGTH CURVES TO ZERO TIME EACH POINT REPRESENTS AVERAGE OF AT LEAST SIX RUNS; STANDARD DEVIATIONS ALWAYS LESS THAN 4%
ELECTRIC STRENGTH (MV/CM)
LOG10 CATHODE NORMALITY

FIG. 10. Dependence of the corrected electric strength of benzene upon cathode normality.

B. The Reversal of Electrode Polarity

The field emission of electrons from negative ions in aqueous solution has been observed under certain circumstances.[12] It is, therefore, necessary to identify the charge carriers causing the effect reported in the previous section.

The mechanism by which charge emission exerts its influence upon the measured electric strength of a dielectric applies equally well whether the emission is of negative charges from the cathode or of positive charges from the anode. It has been shown repeatedly that the measurement of electric strengths of liquids and solids is not influenced by the nature of the anode when metal electrodes are used. While electrons may be emitted from a metallic cathode, apparently positive "holes" are not drawn from a metallic anode in appreciable numbers by the electric field (the emission of a hole into a liquid dielectric is the extraction of an electron from a molecule of the liquid by the anode, leaving a positive ion).

We have, therefore, carried out experiments in which the aqueous electrolyte solution was used as the anode. The results are compared in Fig. 11 with analogous data obtained with the electrolyte as cathode. Except for an apparent small translation of the anode curve, the concentration effect is observed as before. Because of the close correspondence between the two curves, we must assume that emission from the electrolyte surface is possible in both cases. The symmetry of the behavior itself indicates that the charge carrier in both cases is an ion. This conclusion may be supported by simple energy considerations.

Consider the over-all process,

$$Cl^-(Aq) \rightarrow Cl(Aq) + e(\text{benzene}).$$

This may be synthesized from a series of steps, the first of which is the removal of the chloride ion to the

[12] A. Guntherschulze and H. Betz, *Elektrolyt-kondensatoren* (M. Krayn, Berlin, 1937).

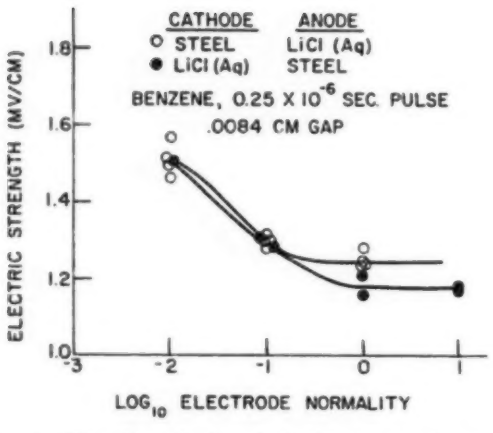


FIG. 11. The effect of reversal of electrode polarity upon the electric strength of benzene.

vacuum:

$$\begin{aligned} Cl^-(Aq) &\rightarrow Cl^-(v) && 3.0 \text{ electron volts,} \\ Cl^-(v) &\rightarrow Cl(v)+e(v) && 3.8 \text{ electron volts,} \\ Cl(v) &\rightarrow Cl(Aq) && -u \text{ electron volts,} \\ e(v) &\rightarrow e(\text{benzene}) && -W \text{ electron volts.} \end{aligned}$$

The positive terms constitute work done on the system. The total energy change is

$$(6.8-W-u) \text{ electron volts.}$$

The quantity u ought to be negligible (if the process is sufficiently irreversible, it will not appear at all). To assess the result, one still ought to have some information about the quantity W; unfortunately, there is none. However, it is sufficient to observe that the predicted electron emission will be less than that from a metal whose vacuum work function is 6.8 volts (provided no specific interaction occurs at the metal-benzene interface). This is because the ion concentration is always less than the concentration of quasi-free electrons in a metal, and because the water-benzene interface has no points such as enhance emission from a metal. The emission of charge actually observed is orders of magnitude greater than that from metals whose vacuum work functions range from 4 to 5 electron volts.

C. The Nature of the Ion

In a separate series of experiments we have investigated the behavior of cathodes containing various negative ions. These experiments were performed with the 0.25-microsecond pulse, using the many-pulse technique, and the results (not corrected for deformation) are reported in Fig. 12. It is clear that the nature of the negative ion has little or no influence on the effect studied here.

There are two conclusions which may be drawn from this result. As an observation supplementary to the discussion of the preceding section, we may say that if electron emission were to occur, the effect of the widely different electron ionization energies of the hydroxyl ion and chloride ion should have been observed (this difference is about 2 electron volts for the gaseous ions). Secondly, as we have had occasion to comment, it is likely that several water molecules accompany the ion in its transfer to the benzene. This will have the result that the effective ion sizes are about the same, a conclusion in harmony with the experimental results.

D. The Single-Pulse Technique

When aqueous electrolyte cathodes of concentration lower than 10^{-2} normal were used in the breakdown studies of benzene, the application of pulses near breakdown invariably resulted in the appearance of a cloud of small particles (1 to 5 microns in diameter) in the gap. Since these particles could not easily be removed between pulses, erroneous breakdown values often resulted. It was, therefore, necessary to use the single-pulse technique with systems involving these low cathode concentrations.

An extension of this technique to systems in which cathodes of higher concentration were used has demonstrated another interesting effect. The results, plotted in Fig. 10, show that the dependence of the electric strength on concentration has disappeared. The single-pulse strength obtained with any cathode whose concentration is greater than 10^{-3} normal is equal to the value obtained from the upper flat portion of the many-pulse curve. Evidently a single pulse is incapable of inducing ion emission. The function of the pulses preceding breakdown, when the many-pulse technique is used, is then to "blur" the Schottky barrier, which is initially too narrow to allow appreciable emission. Although it is impossible to say exactly what the prebreakdown pulses do, it is probable that they create something like an emulsified layer about 100A thick at the interface. In the presence of the electric field, the barrier presented by this layer may be lowered as much as one volt, making intense emission possible. In fact, a group of experiments involving a "double-pulse" technique showed that a single pulse, preceding the breakdown pulse itself, was sufficient to restore the concentration effect almost completely.

The portion of the single-pulse curve in Fig. 10 which lies below 10^{-3} normal in concentration shows a second increase in apparent electric strength. This is not a property of the breakdown system at all, but is caused by the effect of the high electrical resistance of these solutions on the transmission of the short pulses to the breakdown gap.

The appearance of the small, solid particles mentioned earlier deserves further comment. They appeared during pulses near breakdown when very dilute cathodes were used. Under such conditions, the electric field in the gap is homogeneous, so that the entire dielectric, rather than a small region, is subjected to a high field. It is possible that the particles are bits of carbon produced by prebreakdown discharges, and that they show up under the present circumstances because these dis-

charges occur throughout the dielectric rather than in a small region near the anode.

E. The Motion of the Ions

Inasmuch as the duration of the electric field is very short in these experiments, a serious question arises as to whether the ions can really move as rapidly as is required for the validity of our theory. This is to say, the ions must be capable of moving across an appreciable fraction of the gap in a time of the order of a microsecond.

We have no direct knowledge of the mobility in benzene of the ions with which we deal, but we can infer a reasonable value from some measurements which have been made of the transport of ions of another sort in liquid hydrocarbons.[13] When a liquid hydrocarbon is bombarded with x-rays, a variety of ions is formed in it. The mobilities of these ions may be obtained from a measurement of the time interval between their production and their collection on an electrode spaced a known distance away. The ions themselves are both positive and negative, and presumably consist of singly or multiply ionized hydrocarbon molecules or fragments thereof. We have, of course, no information as to the relative sizes of these ions and the ions we use in our experiments. It is probably safe to infer that the mobility in benzene of the inorganic ions plus attached water molecules is comparable with the mobility of ions formed from fairly large hydrocarbon molecules, e.g., *n* hexane. In this case, the mobilities of the several kinds of carriers were grouped about a value of 5×10^{-4} cm^2/v sec.[13]

Using this mobility, we find that an ion may travel 5×10^{-4} centimeter in a microsecond when the electric field is 10^6 v/cm. This distance is about 6 percent of the gap spacing used in our experiments, whereas we require a travel of at least one-third of the gap in order to provide the required distortion of the electric field.

There are two factors, ignored so far, which alleviate this situation. The first is that more than one pulse is necessary to provide the concentration effect. Second and more important, in this estimate we have used a mobility measured at low electric fields. As pointed out earlier, the mobility of ions is expected to increase with electric field at high fields. Evidence that this is so is provided by the fact that the ratio of the upper and lower flat portions of the curve of electric strength *versus* cathode concentration is 1.25 instead of the 1.5 ratio which would result from ohmic behavior. This has a secondary effect which also lessens the requirement of rapid motion of the ions. There is less field distortion; consequently, the distance the ions must travel in order to produce the required distortion is somewhat less. Altogether, it seems that the requirements put on the rapidity of ion motion are not excessive.

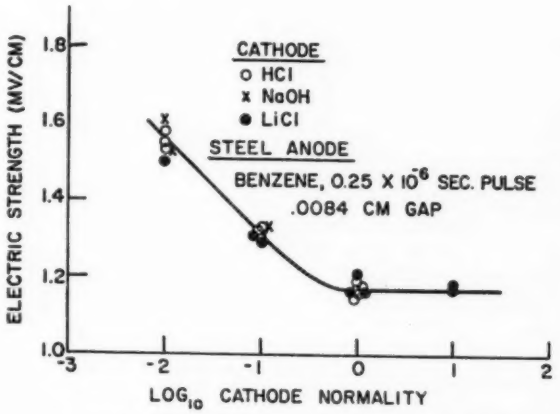


FIG. 12. The effect of the nature of the ion upon the electric strength of benzene.

To support this picture we have substituted a heavy mineral oil for benzene and repeated many of the experiments described in the preceding sections. Stokes's law, which has an approximate validity when applied to the motion of ions through a viscous medium under the influence of an electric field, shows that the mobility of an ion ought to vary inversely with the viscosity of the medium through which it travels. The viscosity of the mineral oil in question is more than a hundred times that of benzene, so that the motion of ions should be inappreciable during a microsecond.

Our experiments on mineral oil show this to be so. There was no observable dependence of the breakdown strength upon cathode concentration, whether the many-pulse or single-pulse technique was used.

ACKNOWLEDGMENT

The authors wish to express their gratitude to Dr. M. H. Hebb for his numerous helpful suggestions and criticisms.

[13] I. Adamczewski, Ann. phys. 8, 309 (1937).

On the Electric Strengths of Aliphatic Hydrocarbons

R. W. Crowe, J. K. Bragg, and A. H. Sharbaugh
General Electric Research Laboratory, Schenectady, New York

(Received August 24, 1953)

Paschen's similarity law, which expresses a certain relationship between the breakdown voltage of a gas, the gas pressure, and the distance between electrodes, is known to fail at high pressures. This is owing, at least in part, to a change in the way in which electrons dissipate energy to the gas molecules. At high enough pressure, the electron transfers energy principally to molecular vibrations. In condensed phases this vibrational barrier is dominant, and one should expect a new kind of similarity law to hold under certain controlled conditions. The liquid aliphatic hydrocarbons provide an illustration of such a law. The electric strengths of nine hydrocarbons of various densities were measured and found to depend linearly upon density. It has also been found that the temperature dependence of the electric strength of a liquid aliphatic hydrocarbon can be accounted for by the change in density of the liquid due to thermal expansion.

I. INTRODUCTION

It was pointed out some fifteen years ago by von Hippel[1] that Paschen's similarity law, which relates the electric strength of a gas to pressure (or density) and electrode separation, cannot hold at very high pressures or for condensed phases. If one extrapolates the curve of breakdown voltage *versus* gas density to densities corresponding to the liquid or solid state, the electric strengths so obtained are of the order of 5×10^7 volts/cm. Experimentally determined values, on the other hand, normally fall between 1×10^6 and 7×10^6 volts/cm for such condensed phases.

The explanation of this discrepancy is fundamental to the interpretation of the electric breakdown of condensed phases. We will therefore discuss the basis for Paschen's law and the reasons for its failure at high density. We will then present evidence which indicates that a new similarity law may hold at high density, the evidence being that such a law is in fact obeyed by the liquid aliphatic hydrocarbons. In some respects, the discussion follows that of reference (1).

II. THE BREAKDOWN OF GASES

A. The Physical Basis of Paschen's Law

It can be shown that a wide variety of breakdown mechanisms all lead to a dependence of the breakdown *voltage*, V, upon gas density, ρ, and electrode separation, δ, in which these variables appear only as the product $\rho\delta$; that is

$$V=f(\rho\delta). \tag{1}$$

This equation is a statement of Paschen's law, a law which was initially established on an empirical basis. One highly idealized model leads to this relationship very simply. Let the breakdown voltage depend only on the electrode separation (δ), on a distance (λ), and on an energy (I) characteristic of the gas. In the following, the quantity λ is assumed to be directly related to the mean free path of an electron. It can then be shown by a dimensional argument that there is a relation

$$F(\lambda/\delta, V\lambda/I\delta)=0 \tag{2}$$

connecting these variables. There are, of course, other choices of the two independent dimensionless ratios which may be made. The fact that λ (which is inversely proportional to ρ) and δ appear only in the combination λ/δ is the basis for Paschen's law for this model.

Experimentally, it turns out that the second ratio used in Eq. (2) is the dominant one for intermediate values of δ/λ, and consequently the breakdown voltage varies nearly *linearly* with δ/λ in this range. However, when δ/λ becomes large there is a deviation from linearity, and if the pressure is large enough (λ small) failure of Paschen's law is observed.

B. The Mechanism of Energy Loss to Gas Molecules

Certain qualitative features of the breakdown mechanism in gases are necessary for our future discussion. Here, as in condensed phases, breakdown occurs when the frequency of ionizing collisions by electrons becomes sufficiently large. The voltage required to bring this about depends upon the way in which the dielectric resists the acceleration of electrons through it.

The retarding barrier in the case of a gas at moderate or low pressure is furnished by inelastic collisions between field-accelerated electrons and gas molecules, which transfer most of the energy accumulated by the electrons into electronic excitation of the molecules. This energy is normally lost to the breakdown process by radiation or by dissociation of the molecule into fragments. Not until a field is reached such that thermal electrons are accelerated to ionizing energy with high probability, can a disruptive avalanche progress through the gas.

As the gas density is increased, however, there is a growing influence of the environment of a molecule upon the structure of its excited electronic states. This influence presumably has the effect of reducing the difference in energy between excited states and ioniza-

[1] A. von Hippel, J. Appl. Phys. 8, 815 (1937).

tion, so much so that electronic excitation in condensed phases is no longer assumed to be an effective barrier to the acceleration of electrons.

III. THE BREAKDOWN OF CONDENSED PHASES

A. The Mechanism of the Process

When the density of a dielectric is sufficiently large, another barrier, which is negligible compared to the excitation barrier at low density, becomes dominant in importance. Indeed, this barrier, furnished by vibrational excitation, grows in importance for the same reason that the other decreases, namely, the interaction between neighboring molecules. The probability of vibrational excitation starts from zero for very low-energy electrons, increases through a maximum, and again decreases for very high-energy electrons whose interaction time is too short to induce appreciable vibration.

In order for breakdown to occur in a condensed dielectric, it is sufficient that the electric field be of such a magnitude that a free electron of any energy below ionization will be accelerated on the average ("low-energy" criterion). This criterion ignores the effect of field-enhanced thermal fluctuations, and consequently overestimates the electric strength; the seriousness of this defect is not known. Nevertheless, electric strengths defined in this way should exhibit the proper dependence on the variables of interest here.

B. A Similarity Law for Condensed Phases

The breakdown mechanism in condensed phases is not really much different than that in the gas phase, only the mechanism of energy loss having changed. It may therefore be useful to explore the possibility that the breakdown voltage of a given liquid or solid is expressible in terms of the same kinds of quantities. In this event, it will involve a characteristic distance of some sort analogous to a mean free path, the separation of electrodes, and a characteristic energy, ϵ, perhaps that at which the rate of loss of energy of an electron is the most severe. We have again the relationship

$$F(\lambda/\delta, V\lambda/\epsilon\delta)=0. \tag{3}$$

Thus, if it were possible to vary λ for a given material, we should find similarity behavior. Unfortunately, this is not so easy as it is with gases, for in condensed phases a change in pressure may be expected to bring about an important change in ϵ as well as in λ.

We may draw another conclusion from this relationship. If the experiments which we analyze are conducted with gaps sufficiently large so that the breakdown *gradient*, V/δ, is independent of δ, then only the second dimensionless quantity of Eq. (3) is important. From this, we can conclude that the breakdown gradient ought to vary linearly with $1/\lambda$, other things being equal. The remainder of this paper is concerned with describing certain experiments in which the electrode spacing was sufficiently large that the breakdown gradient was independent of it. The distance λ was varied in what turned out to be a satisfactory manner, and similarity behavior was observed.

TABLE I. Electric strengths of aliphatic hydrocarbons.

Hydrocarbon		Density	E(MV/cm)	Std. dev. (%)	No. of runs
n-Pentane		0.626	1.21	5.4	10
n-Hexane		0.660	1.29	5.9	11
n-Heptane	25°C	0.679	1.32		3
	50°C	0.658	1.29	5.9	8
	70°C	0.640	1.22	4.4	8
n-Octane		0.707	1.35	5.0	15
n-Decane		0.730	1.42	6.5	19
n-Dodecane		0.751	1.49	6.1	10
n-Hexadecane		0.775	1.50	4.1	9
n-Decane-mineral oil		0.790	1.56	5.1	10
Mineral oil		0.867	1.69	3.1	6

IV. THE NORMAL PARAFFIN HYDROCARBONS

The density of liquid hydrocarbons is known to vary over a considerable range depending upon molecular size and shape. Thus it is possible to vary the density of a liquid dielectric, not by pressure change, but by carrying out experiments on a variety of liquids for which one may hope that the values of the parameters in Eq. (3) other than λ are essentially constant. We report here breakdown experiments on a series of aliphatic hydrocarbons which, in general, show an increase in density as the molecular weight is increased.

A. Experimental Method and Results

The hydrocarbons which were investigated are listed in the first column of Table I; with the exception of mineral oil and the mixture of mineral oil and *n* decane, they are all pure straight-chain alkanes.

The technique of purification of the paraffin hydrocarbons was designed to remove impurities other than paraffin hydrocarbons of different molecular weight. These impurities include sulfur compounds, unsaturated hydrocarbons, and dust particles. The liquids were first washed repeatedly with concentrated H_2SO_4. They were then shaken with mercury for from three to seven days to remove impurities which might later form undesirable films on the surface of the mercury cathode used in the breakdown experiments. This step was followed by filtration through an ultrafine fritted glass filter (pore size 0.9–1.4μ). When possible, the filtering step was preceded by distillation at atmospheric pressure.

The apparatus and procedure involved in the breakdown measurements are described in detail elsewhere.[2,3] All measurements were made using a mechanically polished stainless steel anode and mercury cathode.

[2] Sharbaugh, Crowe, and Bragg, J. Appl. Phys 24, 814 (1953).
[3] Bragg, Sharbaugh, and Crowe, J. Appl. Phys 25, 382 (1954).

Single, constant-length, rectangular voltage pulses of gradually increasing amplitude were applied to the breakdown gap until disruption of the dielectric occurred. The first pulse causing such disruption was recorded as the breakdown voltage of the liquid. Correction for movement of the mercury surface during voltage application was made as described before.[2] The corrected values of the electric strengths are shown in the column of Table I designated $E(MV/\text{cm})$.

B. Dependence of Electric Strength Upon Density

The key to the behavior of the hydrocarbons is apparently contained in the relationship of the distance λ to the density. Let us assume that ϵ (see Eq. 3) is the same for all the hydrocarbons, and that λ may be written as

$$\lambda \sim 1/\sum n_i a_i, \tag{4}$$

where n_i is the number per unit volume of atoms of the ith kind and a_i a constant which has the dimensions of area and represents a sort of scattering cross section appropriate to the jth kind of atom. To be sure, both parts of this assumption are only approximations at best. In particular, the various kinds of atoms which make up the molecules are not distributed entirely at random throughout the dielectric as far as the electron is concerned, and, secondly, differences in chemical binding may exert an influence on both ϵ and λ. Disregarding these effects, we find

$$E = k(n_C a_C + n_H a_H). \tag{5}$$

Since the ratio of carbon atoms to hydrogen atoms is essentially constant for the hydrocarbons we have studied, there should be a linear dependence of the electric strength upon the ordinary density, a plot of which should extrapolate approximately to the origin.

Actually, our experiments show a remarkable agreement with these predictions, better than one would anticipate. In Figs. 1 and 2, the electric strengths of the hydrocarbons are plotted against liquid density.

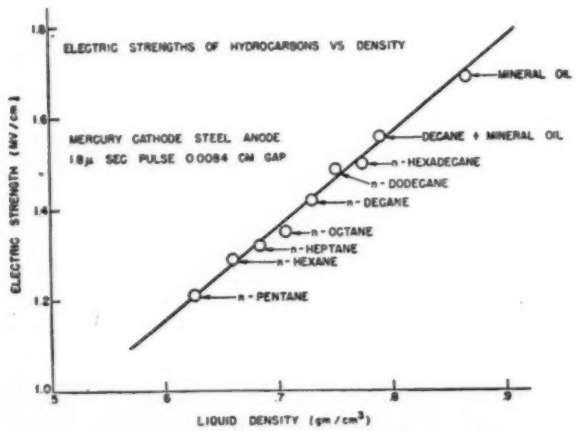


FIG. 1. Density dependence of the electric strengths of aliphatic hydrocarbons.

Within experimental error, the points fall on a straight line.

C. The Question of the Intrinsic Strength

The discussion of the preceding paragraphs has been phrased in terms of properties of the dielectric alone. But it is notorious that measured electric strengths depend on external factors such as the nature of the electrodes, and also upon the presence of exceedingly minute particles. Consequently, no one has ever knowingly measured the "intrinsic strength" of a liquid dielectric, and the present results are to be regarded only as "apparent" electric strengths, characteristic of the method of measurement as well as of the dielectric.

One may expect, however, that the trend shown by these apparent strengths is a characteristic of the properties of the dielectric, although its details may vary with different methods of measurement (electrodes, etc.). In particular, one cannot anticipate that the remarkable agreement of all the results, or the success of the extrapolation to the origin, will be demonstrated by all methods of measurement.

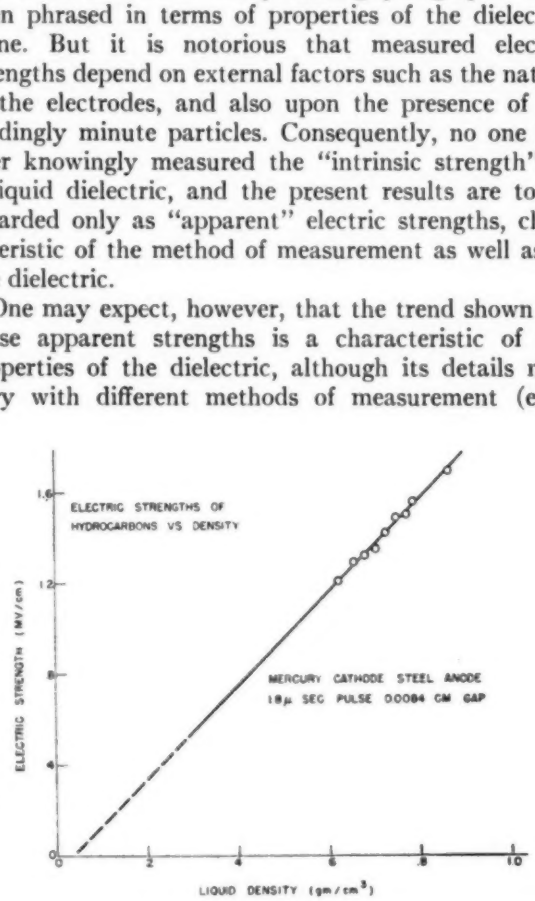

ELECTRIC STRENGTHS OF HYDROCARBONS VS DENSITY
MERCURY CATHODE STEEL ANODE
1.8μ SEC PULSE 0.0084 CM GAP
ELECTRIC STRENGTH (MV/cm)
LIQUID DENSITY (gm/cm³)

FIG. 2. Density dependence of the electric strengths of aliphatic hydrocarbons showing extrapolation.

In connection with this discussion, it seems worth while to mention certain results of other workers. Salvage[4] has reported recently the results of a study of the dielectric breakdown of certain normal alkanes, in which dc voltages were used. Goodwin and MacFadyen[5] have carried out a similar investigation, in which rectangular voltage pulses were employed. Although the results of both investigators show an increase in electric strength with increase in chain length qualitatively like that reported here, no attempt was made to correlate the increase with increase in density of the liquid.

[4] B. Salvage, Proc. Inst. Elec. Engrs. **98**, 227 (1951).
[5] D. W. Goodwin and K. A. MacFadyen, Proc. Phys. Soc. **66**, 85 (1953).

The only recent investigation of the pressure dependence of liquid breakdown is that by Young,[6] who measured the electric strength of carbon dioxide from low pressures through the critical point into the liquid state. He found, as was to be expected, that Paschen's law was obeyed in the gas phase at moderate pressures, but failed at high pressures. However, in the high-pressure gas and liquid region, he observed a linear dependence of electric strength upon density, not unlike that observed in this work.

D. Dependence of Electric Strength upon Temperature

In order to be satisfied with a set of data such as that shown in Figs. 1 and 2, it is necessary to know the temperature dependence of the electric strengths of the liquids. If such a dependence is other than that which would be expected from the normal change of density with temperature, our results have little or no meaning and only accidently give the linear dependence upon density. It is well known that the electric strengths of liquids decrease with temperature as do their densities. In order to demonstrate that a correlation exists, we have measured the temperature dependence of the electric strength of *n* heptane; the results are shown in Fig. 3. The density and electric strength scales in this figure are adjusted so that the relation between a value of the density and the corresponding electric strength is that displayed in Fig. 1. The fact that plots of the density and the electric strength against temperature coincide on this kind of scale shows that the temperature dependence can be interpreted, at least over the indicated temperature range, in terms of changes in liquid density.

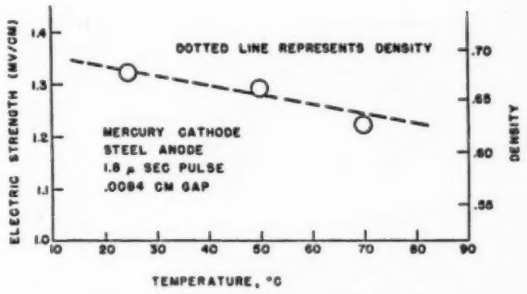


FIG. 3. Temperature dependence of the electric strength and density of *n* heptane.

[6] D. R. Young, J. Appl. Phys. **21**, 222 (1950).

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 3 MARCH, 1954

Normal Reflection of Shock Waves from Moving Boundaries*†

RALPH A. ALPHER AND ROBERT J. RUBIN
Applied Physics Laboratory, The Johns Hopkins University, Silver Spring, Maryland
(Received August 31, 1953)

The problem of the reflection of a shock wave from a moving boundary is examined in an idealized one-dimensional treatment. We consider in detail the head-on collision of an incident step shock wave and a piston moving at constant velocity into a gas at rest. The calculation involves the head-on collision of two shock waves and the subsequent reflection of an altered transmitted wave from the piston. The reflected pressure is calculated for a range of incident shock strengths and piston speeds. For a stationary piston both the absolute and excess reflected pressures are at most eight times the absolute or excess pressure of the incident shock wave (with $\gamma=7/5$). With a moving piston, however, this limiting factor is greater than eight, but finite for finite piston speed and infinite incident shock strength, and is infinite for infinite piston speed and finite incident shock strength. This last result holds even in the limit of the incident wave being an acoustic pulse. The method of solution for shock-wave interaction with a piston receding at constant velocity is also indicated. Although the shock is strengthened on passing through the piston rarefaction wave, the reflected overpressure is less than that resulting from the same initial shock wave reflecting from a stationary piston.

INTRODUCTION

THE normal reflection of shock waves from stationary boundaries has been considered by several investigators.[1] In this paper we consider the problem of the interaction of a shock wave with a piston moving into a gas at rest as a special case of the normal reflection of shock waves from moving boundaries. With a moving piston the gas between the piston and the oncoming shock wave is no longer at rest and the oncoming shock is modified before reaching the reflecting surface. The problem can be handled analytically in the

* This work was supported by the U. S. Bureau of Ordnance, Department of the Navy, under Contract NOrd-7386.
† An account of a portion of this work was presented at the meeting of the Division of Fluid Dynamics, American Physical Society, at Pennsylvania State College, July 1–3, 1953.

[1] See, for example, J. S. Hadamard, *Leçons sur la propagation des ondes* (Librarie Scientifique, A. Hermann, Paris, 1903); S. Chandrasekhar, in a report of limited circulation from Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland, 1943; R. Finkelstein, Phys. Rev. **71**, 42 (1947).

special case of a step shock wave colliding with a piston moving at constant velocity, and this case forms the subject matter of this note. We shall consider only briefly the case of a shock wave impinging on a receding piston.

DESCRIPTION OF SYSTEM

The oncoming shock wave must first interact with a shock wave generated by the piston motion. It has been shown[2] that if a piston has moved at constant velocity V_p for a sufficient time to allow a quasi-steady state to be established, then there is set up before the piston a region of uniform conditions. This region is characterized by an absolute pressure p_p, material velocity $u_p = V_p$, and sound velocity a_p, and is bounded by a shock front whose velocity U_p is determined by the piston speed and the ambient pressure p_0 in the gas at rest ahead of the shock front. The distance between the piston and shock front increases with time according to their relative speed. Since the piston shock wave in this case is a step shock, the problem then is to consider the head-on collision of two unequal step shock waves and the subsequent reflection of the modified oncoming wave from the piston. The complete calculation, therefore, requires only the application of the Rankine-Hugoniot conservation equations. It is assumed that the piston motion is unaltered by the increase in pressure upon reflection of the incident shock wave. Even if the piston motion were altered, the calculations in this paper would still give the instantaneous value of pressure upon reflection at the piston.

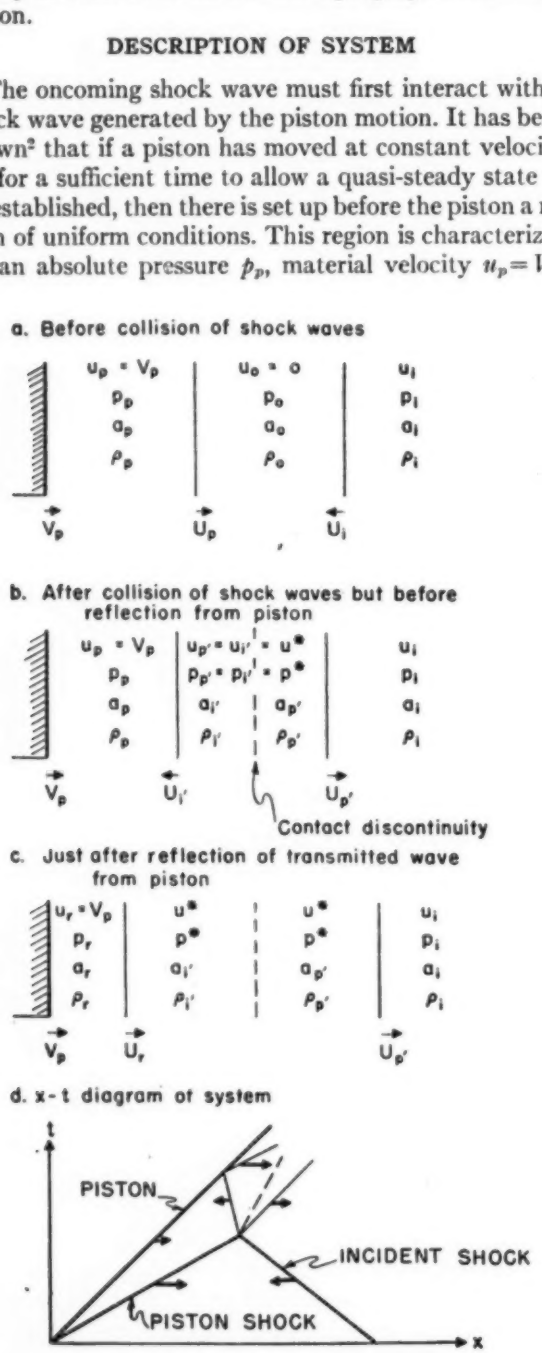


FIG. 1. Sequence of events in the head-on collision of an incident shock wave with a piston moving at constant velocity V_p into a gas at rest. The quantities u, p, a, ρ, and U denote particle velocity, absolute pressure, sound velocity, density, and shock-front velocity respectively.

The sequence of events is illustrated in Fig. 1. The symbols u, p, a, ρ, and U denote particle velocity, absolute pressure, local velocity of sound, density, and shock-front velocity, respectively. The subscripts 0, i, and p refer to ambient conditions in the gas at rest, conditions behind the oncoming step shock wave, and conditions behind the shock wave leading the piston, respectively. Thus Fig. 1a shows the configuration of the system before the collision of the two shock waves. The configuration after the collision is shown in Fig. 1b. Two waves are transmitted, one denoted by subscript i' corresponding to the oncoming wave, and the other denoted by subscript p' corresponding to the wave associated with the piston. The head-on collision of two shock waves was first described by von Neumann.[3] More recently Sauer[4] applied the shock polar method to the numerical description of the altered transmitted waves in such collisions, while Gould[5] has experimentally verified the theoretical description of shock-wave collisions. Upon collision two shock waves are transmitted, between which there is a contact surface which moves at a velocity equal to the mass flow velocity in the region between the shock waves. This surface separates two subregions of different density and sound velocity, although particle velocity and pressure are continuous. The latter continuous quantities are denoted by asterisks in Fig. 1. When the transmitted incident wave is reflected from the piston, the situation is as shown in Fig. 1c. The subscript r denotes quantities behind this reflected wave. Finally, an $x-t$ diagram in Fig. 1d summarizes the sequence of events. It should be

[2] See, for example, R. Courant and K. O. Freidrichs, *Supersonic Flow and Shock Waves* (Interscience Publishers, Inc., New York, 1948), p. 150. See also A. F. Pillow, Proc. Cambridge Phil. Soc. **45**, 558 (1949).

[3] J. von Neumann, Progress Report on the Theory of Shock Waves, National Defense Research Committee, Div. 8, Office of Scientific Research and Development, No. 1140, 1943, p. 15. See also reference 2, p. 177.

[4] R. Sauer, Helv. Phys. Acta **22**, 467 (1949).

[5] D. G. Gould, "The head-on collision of two shock waves and a shock and rarefaction wave in one-dimensional flow." University of Toronto Institute of Aerophysics, Report No. 17 (May, 1952). See also I. I. Glass, "The design of a wave interaction tube," University of Toronto Institute of Aerophysics, Report No. 6 (May, 1950).

noted that we have not considered the subsequent interaction of the reflected wave with the contact surface.

PEAK PRESSURE ON THE PISTON

The altered incident shock wave and piston shock wave can be described by means of the Rankine-Hugoniot conservation relations. The pertinent quantity to be derived is the pressure between the two transmitted shock fronts. Having this pressure, one can then describe the two transmitted waves and calculate directly the reflected pressure. Appropriate coordinate transformations of the usual normal shock relations[6] yields for the shock $U_{i'}$,

$$\left(\frac{u^*+u_p}{a_p}\right)^2=\frac{(1-\mu^2)^2}{1+\mu^2}\frac{[(p^*/p_p)-1]^2}{\mu^2+p^*/p_p}, \tag{1}$$

and for the shock $U_{p'}$,

$$\left(\frac{-u^*+u_i}{a_i}\right)^2=\frac{(1-\mu^2)^2}{1+\mu^2}\frac{[(p^*/p_i)-1]^2}{\mu^2+p^*/p_i}, \tag{2}$$

where

$$\mu^2=(\gamma-1)/(\gamma+1). \tag{3}$$

The quantities V_p and p_i/p_0 are regarded as given. These quantities fix $u_p=V_p$. The pressure p_p and sound velocity a_p in the moving region ahead of the piston are obtained from the following normal shock relations applied to the piston shock front:

$$\left(\frac{u_p}{a_0}\right)^2=\frac{(1-\mu^2)^2}{1+\mu^2}\frac{[(p_p/p_0)-1]^2}{\mu^2+p_p/p_0}, \tag{4}$$

and

$$\left(\frac{a_p}{a_0}\right)^2=\left(\frac{p_p}{p_0}\right)\frac{1+\mu^2p_p/p_0}{\mu^2+p_p/p_0}. \tag{5}$$

Expressions similar to Eqs. (4) and (5) with the subscript i instead of p yield u_i and a_i. Since one is interested in p^*, we add the square roots[7] of Eqs. (1) and (2), divide through by a_0 and replace a_p/a_0 and a_i/a_0 by the square root of Eq. (5) and its equivalent for a_i/a_0. The result can be written in the following form:

$$\frac{(1+\mu^2)^{\frac{1}{2}}}{1-\mu^2}\frac{V_p}{a_0}+\frac{y_i-1}{(\mu^2+y_i)^{\frac{1}{2}}}=\left(\frac{1+\mu^2y_i}{\mu^2+y_i}\right)^{\frac{1}{2}}\frac{y^*-y_i}{(\mu^2y_i+y^*)^{\frac{1}{2}}}+\left(\frac{1+\mu^2y_p}{\mu^2+y_p}\right)^{\frac{1}{2}}\frac{y^*-y_p}{(\mu^2y_p+y^*)^{\frac{1}{2}}}, \tag{6}$$

where

$$y^*=p^*/p_0,\quad y_i=p_i/p_0,\quad \text{and}\quad y_p=p_p/p_0.$$

[6] See reference 2, p. 116 and following.

[7] The same sign (in this case the plus sign) is used for the two terms to yield a physically reasonable result in the case of equal shocks $p_p=p_i$.

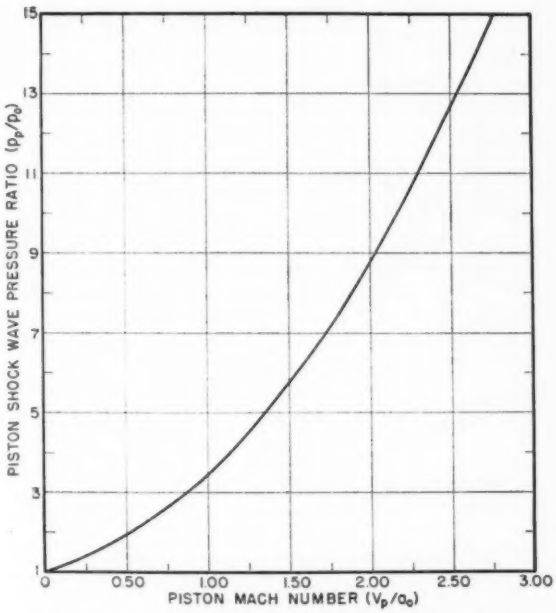


FIG. 2. Piston shock-wave pressure ratio as a function of piston Mach number, for $\gamma=7/5$.

Equation (6) can be solved graphically for y^* in terms of y_p, y_i, and V_p/a_0, which are taken to be known.[8] Having values of y^*, one can then calculate the absolute pressure p_r acting on the piston behind the reflected shock, from the following relation:[9]

$$y_r=p_r/p^*=\frac{(2\mu^2+1)(y^*/y_p)-\mu^2}{1+\mu^2y^*/y_p}. \tag{7}$$

The quantity $y_p(=p_p/p_0)$ is shown in Fig. 2 for a selected range of piston speeds V_p/a_0, in the case $\gamma=7/5$.

There are a variety of ways of presenting the results of these calculations, two of which are given below. In Fig. 3 we have plotted p_r/p_p *versus* $y_i(=p_i/p_0)$ for the several piston speeds considered, again for $\gamma=7/5$. This quantity is the factor by which the absolute pressure, acting on the front of the piston, is increased upon reflection of the incident shock wave. It should be noted that for a given y_i, p_r/p_p decreases with increasing piston speed; however, it can be seen from Fig. 2 that p_p increases with piston speed more rapidly than does p_r. The enhancement of the reflected pressure due to the piston motion is better illustrated by Fig. 4, in which we have plotted P_r/P_i *versus* $y_i(=p_i/p_0)$. The quantity P_r is the excess of the reflected pressure over that which had acted on the piston, and can be shown to be

$$P_r=(y_ry^*-y_p)p_0, \tag{8}$$

while $P_i(=p_i-p_0)$ is the excess pressure in the oncoming shock wave.

[8] Gould, (see reference 5), has solved an equivalent of Eq. (6) for y^* by iteration in the case $y_p=5.33$, $\gamma=1.4$, for a range of incident shock strengths $1\leq y_i\leq 29$.

[9] See reference 2, p. 153.

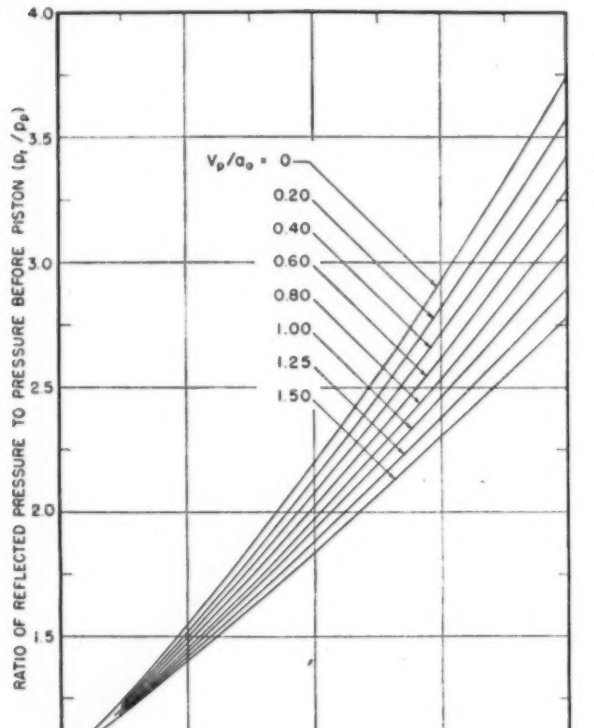


FIG. 3. Ratio of the absolute shock-wave pressure after reflection (p_r) to the absolute pressure acting on the moving piston before reflection (p_p) as a function of the incident shock strength (p_i/p_0) for a range of piston velocities, with $\gamma=7/5$.

LIMITING CASES

It is of some interest to consider various limiting cases in connection with the aforementioned results. If we examine Eq. (7), it is easy to see that for $V_p/a_0=0$, (i.e., $y^*=y_i$, $y_p=y_0$) the quantity P_r/P_i approaches a value of eight as y_i increases indefinitely. For each piston speed there is an asymptotic value approached by P_r/P_i as y_i increases, this value increasing with increasing piston speed. These asymptotic values can be calculated from a limiting form of Eq. (6). For large y_i, i.e., for

$$y_i \gg \frac{2\mu^2(\mu^2+1)}{(\mu^2-1)^2}\left(\frac{V_p}{a_0}\right)^2, \tag{9}$$

Eq. (6) can be reduced to

$$1-\left(\frac{1+\mu^2 y_p}{\mu^2+y_p}\,\frac{y^*}{y_i}\right)^{\frac{1}{2}}=\mu\frac{(y^*/y_i)-1}{(\mu^2+y^*/y_i)^{\frac{1}{2}}}. \tag{10}$$

TABLE I. Limiting excess reflected pressure for large y_i.

$V_p/a_0=$	0	0.2	0.4	0.6	0.8	1.0	1.25	1.50
$P_r/P_i=$	8	8.93	9.88	10.5	11.7	12.5	13.5	14.3

Equation (10) can be solved for y^*/y_i for selected y_p and, since y_i is large, one can then compute from Eq. (7) that

$$\frac{P_r}{P_i}\cong\frac{1+2\mu^2}{\mu^2}\,\frac{y^*}{y_i}=\frac{8y^*}{y_i} \quad \text{for} \quad \gamma=7/5. \tag{11}$$

The limiting values of P_r/P_i approached by the curves plotted in Fig. 4 as y_i increases indefinitely are given in Table I.

These then are the asymptotic values of P_r/P_i for an infinitely strong shock reflected from a piston moving at finite speed. A different situation obtains in the case of an oncoming shock wave of finite strength reflected from a very rapidly moving piston. Since Eq. (6) is symmetrical in y_i and y_p, one can readily find limiting values of

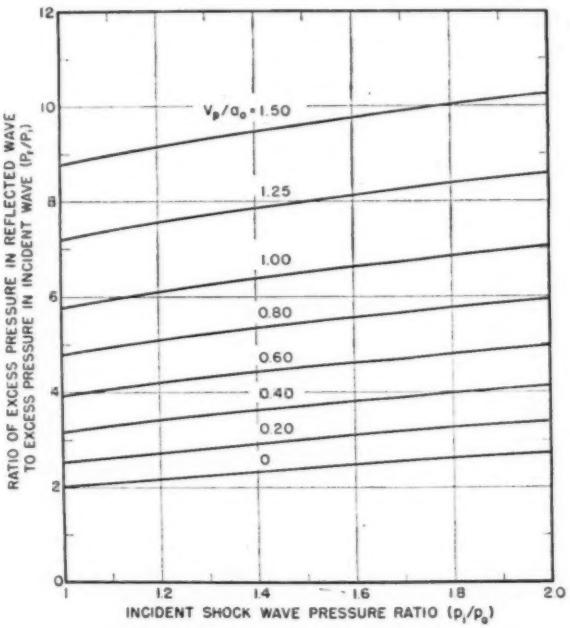


FIG. 4. Ratio of the excess P_r of the reflected shock-wave pressure over that which had acted on the piston to the excess, P_i, of the incident shock-wave pressure over ambient, as a function of $y_i=1+P_i/p_0$, for a range of piston velocities, with $\gamma=7/5$.

y^*/y_p as y_p becomes large with y_i constant. In this case one finds that P_r/P_i becomes infinite in the limit of infinite y_p.

Another limiting case which is of interest follows from Eqs. (6) and (7) for equal strength of the incident shock and piston shock waves, i.e., $y_i=y_p$. The initial collision of the two fronts is equivalent to the reflection of one front from a stationary piston, while the subsequent reflection of the transmitted wave is given by Eq. (7). For very strong shock waves, with $\gamma=7/5$, one obtains a factor of enhancement of 36 from Eq. (7).

Finally, the calculation of pressure enhancement upon reflection in the limiting case of an oncoming acoustic pulse requires that one consider the trans-

mission of such a pulse through a shock wave. This latter problem has been considered by a number of investigators.[10] The required result follows directly from Eq. (6). For an acoustic pulse we replace y_i by $y_i=1+\epsilon$. Since ϵ is small, then y^* will differ from y_p by a quantity of the same order, *viz.*, $y^*=y_p+\delta$. If one replaces y_i and y^* in Eq. (6) by the foregoing expressions, and retains only terms to first order in ϵ and δ, it follows that

$$\frac{\delta}{\epsilon}=\frac{1+(1-\mu^2)(y_p-1)/[2(1+\mu^2)]+\mu^2(y_p-1)/[2(\mu^2+y_p)]+[(\mu^2+y_p)/(1+\mu^2)]^{\frac{1}{2}}}{1-(y_p-1)/[2(\mu^2+y_p)]+[(1+\mu^2y_p)/[y_p(1+\mu^2)]}. \tag{12}$$

From the definition of δ and ϵ we see that the quantity δ is the amplitude of the transmitted acoustic pulse corresponding to an incident acoustic pulse of amplitude ϵ, both in units of p_0. The ratio δ/ϵ, computed from Eq. (12) is plotted *versus* y_p in Fig. 5 for the case $\gamma=7/5$. For large y_p, the ratio δ/ϵ is given by

$$\frac{\delta}{\epsilon}\cong\frac{(1-\mu^2)y_p}{1+\mu^2+2\mu(1+\mu^2)^{\frac{1}{2}}}. \tag{13}$$

In considering the reflection of the transmitted acoustic pulse, one can still treat the transmitted pressure in the acoustic approximation compared to the ambient pressure before the piston, p_p. Hence, the absolute reflected pressure will be $2\delta p_0$ and the over-all amplification of the incident acoustic pulse amplitude due to reflection will be $2\delta/\epsilon$. These considerations, together with Eq. (12), have been employed in obtaining points plotted at $y_i=1$ in Fig. 4. It may be noted that the curves in Fig. 4 are rather flat, so that as a crude approximation one may use the limiting acoustic case to describe the transmission and reflection of weak shocks.

RECEDING PISTON

It is not possible to treat the case of a shock wave interacting with a piston receding at constant velocity in as simple a manner as the head-on collision problem. However, one can utilize studies that have been made of the collision of rarefaction waves and shock waves[5,11,12] to draw some general conclusions about the interaction. One result is that the transmitted shock wave is accelerated, although the overpressure upon reflection will be less than that for the reflection of the same shock wave from a stationary piston.

From Nicholl's work[12] one can, as an example, calculate the reflected pressure acting on a piston receding at Mach numbers in the range from 0.23 to 0.55 and compare the result with the stationary case. For Mach number 0.23, with an incident shock strength of $y_i=2.48$ the excess reflected pressure is $3.58p_0$ for the receding piston and $4.51p_0$ for a stationary piston. For Mach number 0.55, with an incident shock strength of $y_i=2.92$, the excess reflected pressure is $4.07p_0$ for the receding piston and $6.32p_0$ for a stationary piston.

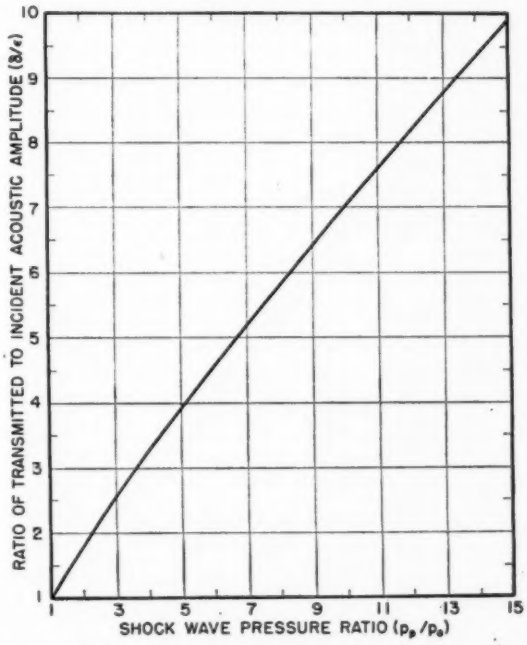


FIG. 5. Ratio of the transmitted to incident overpressure, δ/ϵ, for acoustic pulses transmitted through shock waves, as a function of shock-wave pressure ratio p_p/p_0 for $\gamma=7/5$.

ACKNOWLEDGMENTS

The authors would like to thank Dr. R. C. Herman for helpful comments on the manuscript, Miss Doris Rubenfeld for help with the illustrations, and Mrs. Carol Oberlin for typing the manuscript.

[10] See, for example, J. M. Burgers, Proc. Koninkl. Ned. Akad. Wetenschap. **49**, 273 (1946); D. Blokhintzev, J. Acoust. Soc. Am. **18**, 329 (1946); G. F. Carrier, Quart. Appl. Math. **6**, 367 (1949); F. K. Moore, "Unsteady oblique interaction of a shock wave with a plane disturbance," National Advisory Committee for Aeronautics, Technical Note 2879 (January, 1953). This last work contains some additional pertinent references. Burgers' result [Eq. (22) in his paper] is equivalent to our Eq. (12), except for an error in sign, while Blokhintzev's work contains so many typographical errors as to make it extremely difficult to use.

[11] H. E. Moses, J. Appl. Phys. **19**, 383 (1948). The collision of a shock wave with the rarefaction wave produced by an accelerated piston is treated in this work.

[12] C. I. H. Nicholl, "The head-on collision of a shock and rarefaction waves," University of Toronto Institute of Aerophysics, Report No. 10 (October, 1951). Nicholl produced rarefaction waves by the bursting of a diaphragm in a shock tube. The wave thus produced can be regarded as having been generated by a piston located at and moving with the constant velocity of the contact surface in the shock-tube expansion chamber.

Heat Transfer Measurements at Sodium–Stainless Steel Interface

JAMES W. MOYER AND WILLIAM A. RIEMEN
Knolls Atomic Power Laboratory, Schenectady, New York*
(Received August 24, 1953)

Experiments to measure the heat-transfer coefficient at a sodium type-347 stainless steel interface are described. The temperature range is 100–500°C. It is concluded that wetting does not play a significant role in the heat transfer.

INTRODUCTION

EXPERIENCE with mercury boilers has shown that a critical change in the heat-transfer coefficient occurs as the mercury wets the metal container. There has been a general feeling that a considerable thermal barrier may exist at the interface of stainless steel and liquid sodium when the angle of contact is large.

Accordingly, apparatus was designed to permit measurement of the heat transfer coefficient at the interface as a function of the temperature of the liquid sodium from 100 to 500°C. Clean surfaces (degreased and sometimes polished) were used, outgassed thoroughly under high-vacuum conditions, and the contact made with triply distilled sodium. In two of the experimental arrangements it was possible to make visual observations of the progress of wetting. Type-347 stainless steel was used throughout.

METHOD

In two of the experimental arrangements heat flowed through a uniform conductor where a flux measurement could be made, then through an interface into a reservoir of sodium, schematically shown in Fig. 1.

The temperature distribution of the conductor was continuously recorded. A change in the thermal impedance of the interface I should cause a change in the temperature distribution, provided no other conditions are changed.

In the third experimental arrangement the situation was reversed. The sodium was the source of heat, and the heat passed across the interface and through a uniform cylinder, the far end of which was cooled. The flux was again measured in the uniform conductor.

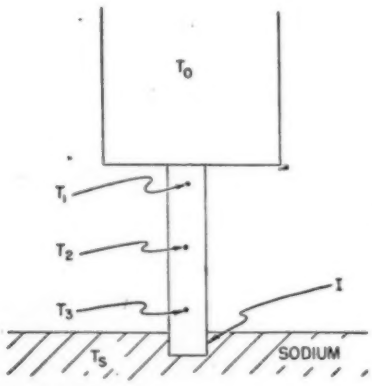


FIG. 1. Heat-flow path and temperature-measurement points.

Experimentally, Q, the heat flow at equilibrium, was determined by the flux-measuring thermocouples T_1 and T_2, or T_2 and T_3:

$$Q=KA\frac{T_1-T_2}{X_2-X_1}=KA\frac{T_2-T_3}{X_3-X_2}, \tag{1}$$

where A is the cross-sectional area of the uniform conductor and K its thermal conductivity.

If the film impedance is not negligible, and if temperature gradients in the sodium are neglected, then

$$T_3-T_s=Q\left(\frac{L-X_3}{KA}+\frac{1}{UA'}\right), \tag{2}$$

where U is the heat-transfer coefficient of the interfacial film, A' the interfacial area, and L the conductor length. The coefficient is given by

$$\frac{1}{U}=A'\left[\frac{(T_3-T_s)}{(T_2-T_3)}\frac{(X_3-X_2)}{KA}-\frac{L-X_3}{KA}\right]. \tag{3}$$

The experiments were designed to measure the ratio $(T_3-T_s)/(T_2-T_3)$ as precisely as possible.

EXPERIMENT I

A sketch of the apparatus for the first attempt is shown in Fig. 2. The entire thermal path was a stainless steel blade $\frac{1}{16}$ in. thick, 1 in. long, and $\frac{1}{2}$ in. wide. The flux in the specimen ranged between 30 and 50 w/sq cm. The specimen was immersed about $\frac{3}{16}$ in. Since the conductivity of the sodium is roughly four times that of stainless steel, the heat passed from the specimen into the sodium in about the first $\frac{1}{16}$ in. of the immersed section. Thus, the heat flux at the interface was approximately half that in the specimen itself.

The apparatus was carefully outgassed in high vacuum before triply-distilled sodium was admitted. By means of an iron slug and an external rotating magnet, the sodium was stirred.

Radiation shields around the specimen cut down the lateral heat loss from the specimen by about 80 percent.

The first two runs with this apparatus yielded the only indication of other than negligible thermal impedance of the interface. At about 375°C, after immer-

* Operated by the General Electric Company for the U. S. Atomic Energy Commission.

sion at lower temperatures from 3 to 4 hours, the temperature distribution suddenly changed to indicate a large increase in the heat-transfer coefficient. From the change in temperature distribution, it was deduced that the heat-transfer coefficient of the film was about equal to that of $\frac{1}{4}$ in. of stainless steel before the change took place (this value is about 0.41 w/sq. cm-deg C). After the change the heat-transfer coefficient was so large that for all practical purposes the film did not exist. The experimental error was such that the minimum film impedance detectable would correspond to a coefficient of about 4 w/sq cm-deg C, which is equivalent to the impedance of about 0.025-in. stainless steel.

Several attempts were made to reproduce these data on subsequent tests but without success. All other tests indicated no change in the temperature distribution, even after wetting occurred.

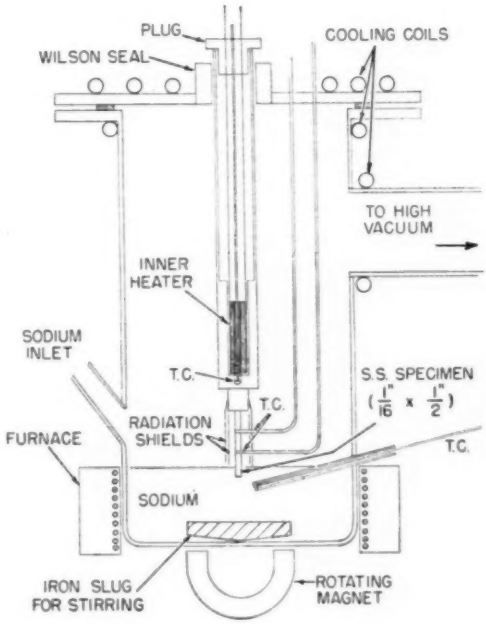


FIG. 2. First experimental arrangement.

EXPERIMENT II

This apparatus is shown in Fig. 3. It was designed to increase the interfacial area, reduce the lateral heat loss from the specimen, and increase the heat flux through the interface. Heat fluxes between 100 and 200 w/sq cm were maintained. Data were taken up to 500°C. The compensating heaters were adjusted to produce a zero temperature gradient in the walls of the supporting "sphere" nearest the specimen.

The chief uncertainty in measurements made with this apparatus came from inability to keep constant the amount of sample immersed. Arrangements were made to adjust the level of the sodium surface by means of an auxiliary sodium reservoir which could be raised and

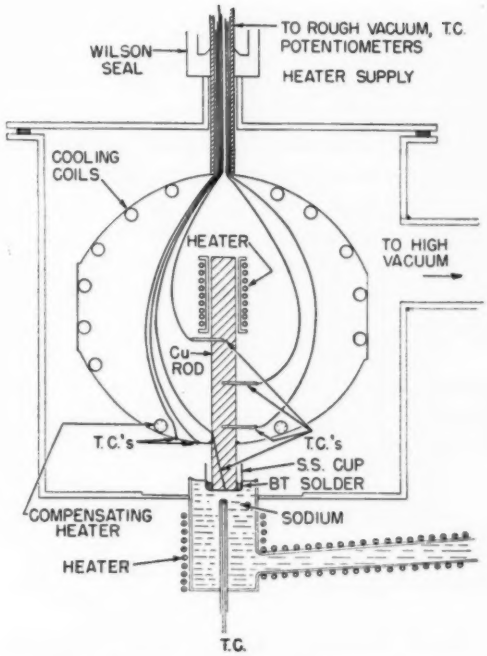


FIG. 3. Second experimental arrangement.

lowered. In spite of this device, the wetting did not take place uniformly around the specimen, so that more contact was made in regions completely wet than in other regions.

Runs were made with highly polished stainless steel cups and also with highly oxidized sufraces. No experiment yielded a measurable amount of thermal impedance of the film. The minimum film impedance in this setup that could be measured was equivalent to about 0.020 in. of stainless steel.

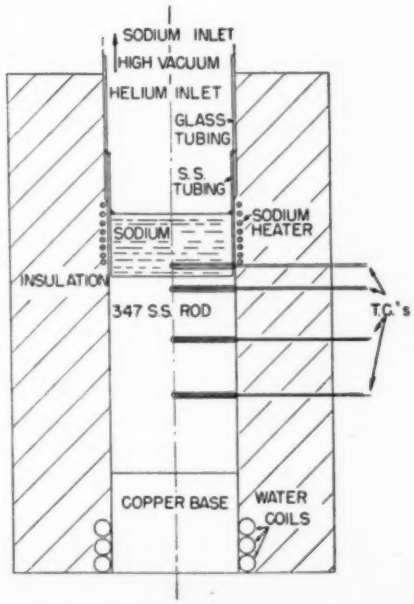


FIG. 4. Third experimental arrangement.

EXPERIMENT III

Experiment III was a low-flux experiment, but one which gave no difficulties due to variable area of contact or heat losses. The arrangement is shown in Fig. 4. Fluxes were of the order of 7 w/sq cm. In this apparatus helium could be admitted over the sodium surface at all times to reduce evaporation.

The data indicated that no thermal impedance existed larger than that equivalent to about 0.010 in. of stainless steel, which was the limit of precision of the measurement.

CONCLUSIONS

A general weakness of this experiment which is common to many other experiments involving stainless-steel surfaces is the lack of characterization of the surface. It is well known that the thickness and composition of oxide layers on stainless steel change markedly with heat treatment, polishing or abrasion, type of machining, and treatment by solutions. It is probably true that in no two experiments were the stainless-steel surfaces identical. Even with the same sample the cleaning treatment between runs undoubtedly affected the surface conditions.

The only evidence of thermal impedance of the interface came with preliminary measurements in the first apparatus. In view of the lack of substantiation of the earlier observed effect, one is forced to conclude that under the conditions of these experiments wetting plays little, if any, part in the heat transfer at the sodium-stainless steel interface.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 3 MARCH, 1954

Low-Temperature Internal Friction and Elasticity Effects in Vitreous Silica

M. E. Fine, H. Van Duyne, and Nancy T. Kenney
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received August 17, 1953)

Near 35°K the internal friction of vitreous silica reaches a peak value which increases with frequency. The temperature curves of the torsion and Young's moduli are concave upward with minima near 65°K and are independent of frequency. These effects are reversible with temperature. Their cause cannot be the usual type of relaxation process or anything involving diffusion or large amounts of shear.

IN vitreous silica attenuation (db loss per unit length) of 10 mc/sec acoustic waves increases sharply on cooling from room temperature to 63°K (McSkimin[1]). Furthermore, the values of attenuation[1] at 1.6 and 4.3°K are considerably lower than at 63°K suggesting the presence of a peak at some temperature between. The attenuation was found to increase almost linearly with frequency,[1] evidence according to McSkimin,[1] that the attenuation is by absorption in the material (internal friction) rather than scattering of acoustic waves by inhomogeneities.

Attenuation of 37 kc/sec acoustic waves (Marx and Sivertsen[2]) also increases on cooling to 77°K, but the relative increase is much less than for the higher-frequency acoustic waves.

In the present investigation Young's modulus (E) the torsion modulus (G), and the internal friction ($1/Q$ the fractional energy absorbed per radian) of vitreous silica (Amersil, clear and strain free) were determined as functions of temperature from 2 to 200°K in the frequency range 50 to 375 kc/sec using the resonant piezoelectric method.[3,4] For each series of measurements the sample was cemented to a quartz crystal with Thiokol cement for the low temperatures and with red sealing wax for the higher temperatures. The sample and crystal were matched in length so that the joint occurred at a stress node. For longitudinal vibrations both $5°X$ and $-18°X$ cut quartz crystals were used; for torsional vibrations a cylindrical quartz crystal with length along the X axis was used. These crystals were calibrated over the temperature range of the investigation to enable calculation of the properties of vitreous silica from the experimental measurements.

The cryostat had a standard double Dewar arrangement. The outer Dewar flask contained liquid nitrogen for cooling to 77°K. The inner Dewar flask contained liquid helium for cooling below 77°K. The piezoelectric apparatus was in the helium Dewar flask and in an evacuated chamber into which helium gas was expanded to a pressure of several mm of Hg. The working chamber contained a heating coil for controlling the temperature above 4.2°K. Temperatures below 20°K were measured with a carbon resistance thermometer, above 20°K with a copper-constantan thermocouple.

In calculations the expansion data of Souder and Hidnert[5] for vitreous silica and the measured density of 2.21 grams/cm were used.

[1] H. J. McSkimin, J. Appl. Phys. 24 (1953).
[2] J. W. Marx and J. M. Sivertsen, J. Appl. Phys. 24, 81 (1953).
[3] F. C. Rose, Phys. Rev. 49, 50 (1936).
[4] W. P. Mason, *Piezoelectric Crystals and Their Application to Ultrasonics* (D. Van Nostrand and Company, Inc., New York, 1950), pp. 83, 90, 151.
[5] W. Souder and P. Hidnert, J. Research Natl. Bur. Standards 21, 19 (1926).

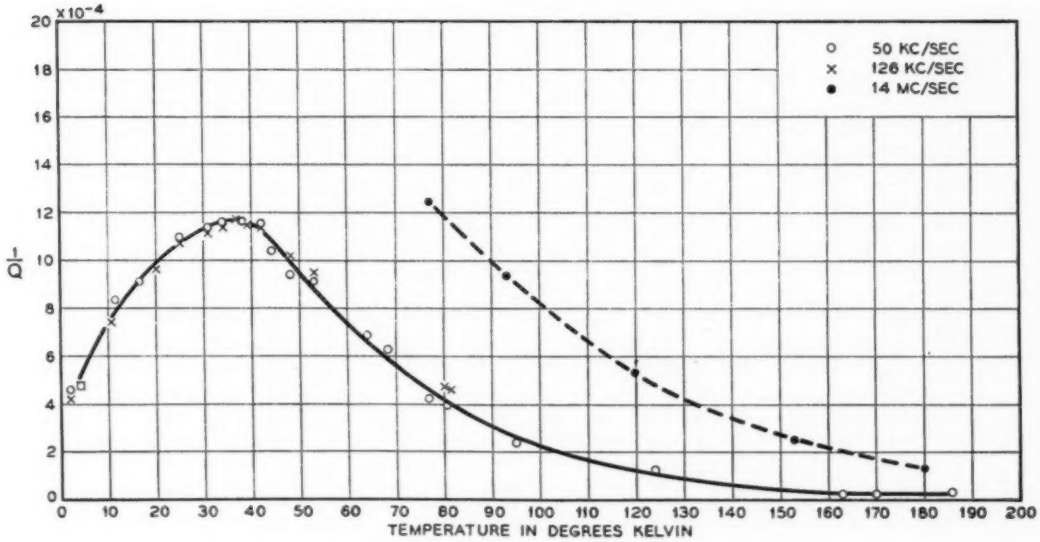


FIG. 1. Internal friction ($1/Q$) of vitreous silica for longitudinal waves. The data at 14 mc/sec are from McSkimin (see reference 1). The samples for the 50 and 126 kc/sec measurements were 5.1 mm in diameter and 34.5 mm long and 22.4 mm long, respectively.

INTERNAL FRICTION

The internal friction of vitreous silica for longitudinal waves, shown in Fig. 1, rises to a peak near 35°K. The temperature of maximum absorption and the internal friction at the maximum are approximately the same for longitudinal waves of 50 and 126 kc/sec. Measurements, not shown, were made at 375 kc/sec; absolute values of internal friction were not obtained but the absorption peak was found to be near 35°K.

The internal friction of vitreous silica for torsional waves, Fig. 2, also rises to a peak near 35°K. The peak value, somewhat lower than for longitudinal waves, increases slightly with increase of frequency from 66 to 201 kc/sec. The internal friction peak for 10 mc/sec shear waves,[1] Fig. 2, appears to be near 35°K, but the peak value must be considerably higher than for the lower frequencies.

YOUNG'S MODULUS, SHEAR MODULUS, POISSON'S RATIO, AND COMPRESSIBILITY

The Young's modulus-temperature and the shear modulus-temperature curves are concave upward, a broad minimum occurring for each near 65°K, Figs. 3 and 4. There seem to be some variations among modulus values from sample to sample. The variation, considerably larger than the experimental error, does not corre-

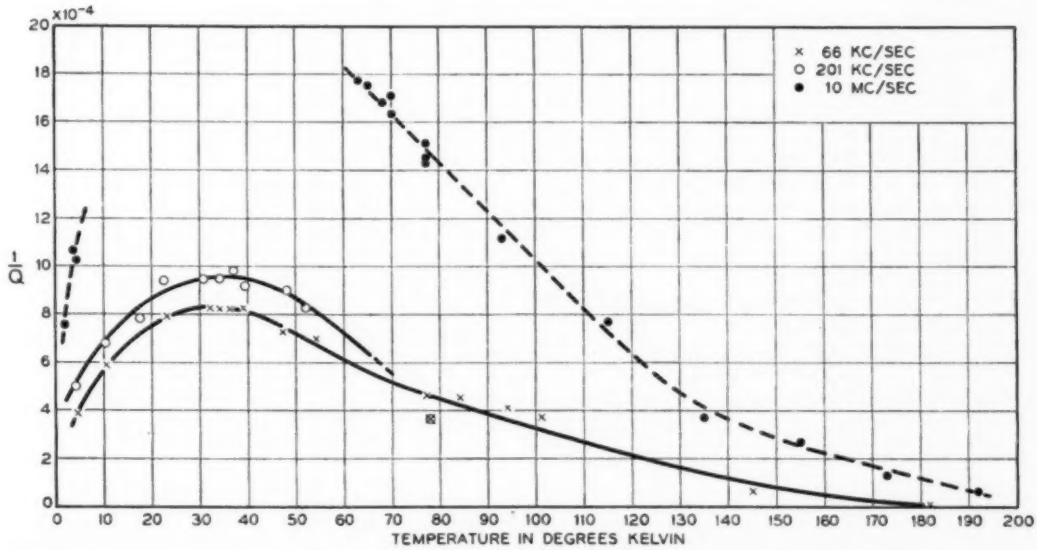


FIG. 2. Internal friction ($1/Q$) of vitreous silica for torsional waves. The data at 10 mc/sec are from McSkimin (see reference 1), and refers to shear waves. The sample for the lower-frequency measurements was 5.1 mm in diameter and 28.1 mm long.

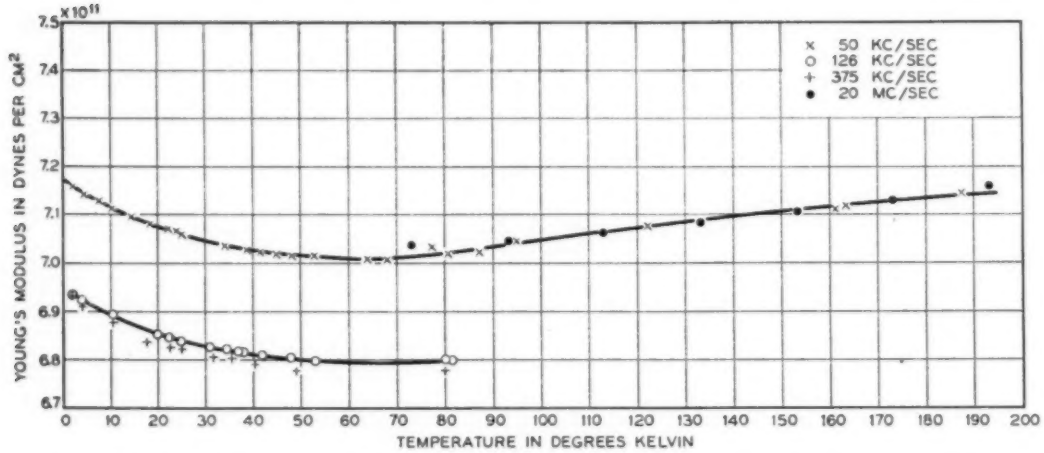


FIG. 3. Young's modulus (E) of vitreous silica. The data at 20 mc/sec are from McSkimin (see reference 1). The samples for the lower-frequency measurements were 5.1 mm in diameter and 34.5 mm long for 50 kc/sec and 22.4 mm long for 126 and 375 kc/sec.

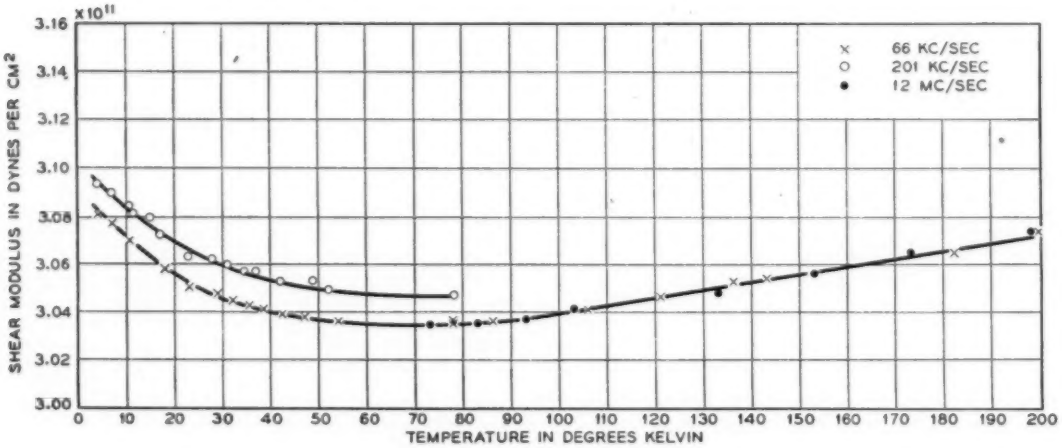


FIG. 4. Torsion modulus (G) of vitreous silica. The data at 12 mc/sec are from McSkimin and refers to shear waves. The sample for the 66 and 201 kc/sec measurements was 5.1 mm in diameter and 28.1 mm long.

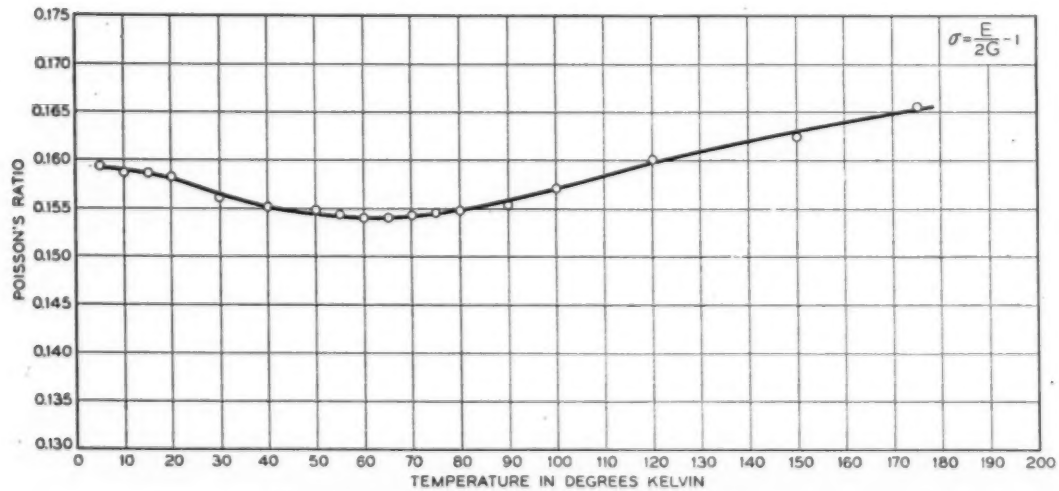


FIG. 5. Poisson's ratio (σ) for vitreous silica. $\sigma = E/2G - 1$.

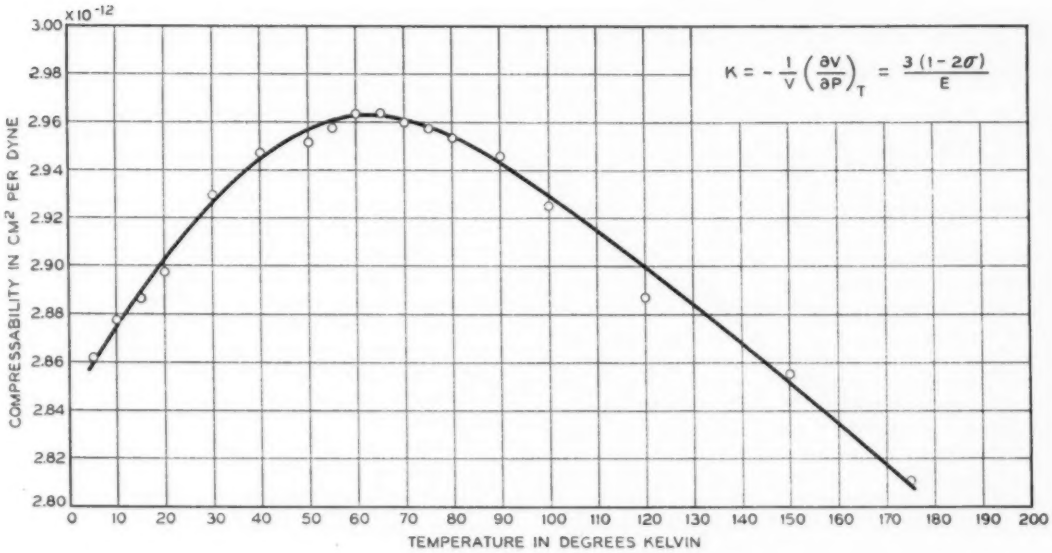


FIG. 6. Compressibility (K) of vitreous silica

$$K=-\frac{1}{V}\frac{(\partial V)}{(\partial P)_T}=\frac{3(1-2\sigma)}{E}.$$

late with frequency. However, all of the $E-T$ curves and all of the $G-T$ curves are respectively parallel.

In vitreous silica Poisson's ratio, Fig. 5, is unusually low (0.15); that is, the lateral contraction accompanying a normal stress is small.

The compressibility (K) mirrors the moduli curves as shown in Fig. 6.

DISCUSSION AND CONCLUSIONS

The usual theoretical treatment of glasses[6,7] supposes that the interatomic forces are similar to those in crystalline solids. In normal crystalline solids E and G increase and K decreases on cooling approaching constant values for very low temperatures. In vitreous silica the absorption peak and the temperature variation of E, G, and K are deviations from normal elastic behavior. (The density of vitreous silica varies abnormally with temperature; the density appears to be maximum near 190°K.[5])

The nature of these effects imposes some limitations as to the mechanism one might propose for their cause. In the first place a normal-type relaxation process is eliminated since the height of the absorption peak increases with frequency and also the modulus-temperature curves differ in some respects from those for a relaxation effect.[8] (In a relaxation process the modulus is expected to decrease from the unrelaxed modulus to the relaxed modulus on heating through the temperature range of the absorption peak; like the absorption peak, in a relaxation process this modulus change occurs at higher temperatures as the frequency is increased.) The reversibility of E, G, and $1/Q$ with temperature and the low temperatures involved make improbable any transformation involving diffusion or large shear motions of the ions.

In vitreous silica as well as in the crystalline forms the building blocks are usually considered to be SiO_4 tetrahedra.[7] The bond angles and interatomic distances are slightly different in each of the crystalline forms; in vitreous silica there is presumably a distribution of bond angles and interatomic distances about the most probable values. The low-temperature effects in vitreous silica, it seems likely, involve this randomness since, as shown by our calibration measurements, there is no absorption peak for acoustic energy in α quartz. The sort of thing that comes to mind is some rapid increase with temperature in the modes of thermal vibration. In this respect the heat capacities of vitreous silica[9,10] at low temperatures are higher than the heat capacities of quartz[9] or cristobalite.[10] Thus there are extra "sinks" for thermal energy in vitreous silica. The absorption of acoustical energy perhaps occurs through interaction with these.

ACKNOWLEDGMENTS

We are indebted to H. E. Johnson for furnishing the density of our sample rods and to O. L. Anderson, W. O. Baker, W. C. Ellis, W. P. Mason, H. J. McSkimin, and E. E. Schumacher for helpful discussion and encouragement.

[6] W. H. Zachariasen, J. Am. Chem. Soc. **54**, 3841 (1932).
[7] B. E. Warren, J. Appl. Phys. **8**, 645 (1937).
[8] See for example C. Zener, *Elasticity and Anelasticity* (University of Chicago Press, Chicago, 1948), p. 48.
[9] W. Nernst, Ann. Physik. **36**, 421 (1911).
[10] F. Simon, Ann. Physik. **68**, 241 (1922).

Letters to the Editor

Shot Dependence of *p-n* Junction Phototransistor Noise

A. Slocum and J. N. Shive
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received November 23, 1953)

SEMICONDUCTOR noise has been the subject of considerable study since the advent of the transistor. Early measurements[1] of the noise figures of point contact transistors yielded values around 50 db when measured in the 1–5 kc range, and showed an inverse dependence of noise figure upon measuring frequency. Measurements of noise in germanium filaments[2] have confirmed the $1/f$ relationship and pointed to a dependence of noise power on dc bias field. More recently, the noise characteristics of certain *n-p-n* junction transistors[3] have been found to agree closely with the shot noise relationship for measuring frequencies of a few kilocycles. Noise in single *p-n* junctions[3] showed a dependence upon dc bias current and an inverse dependence on frequency.

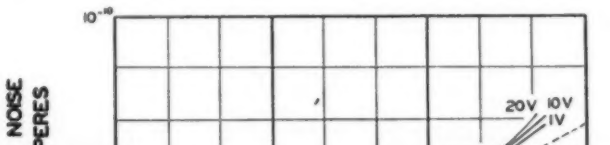


FIG. 1. *p-n* junction current noise. Can be made to approximate shot noise.

Early measurements on *p-n* junction phototransistors[4] indicated noise one or two orders of magnitude in excess of shot noise, with little dependence on the photocurrent. Lately, however, more systematic investigations of noise in experimental *p-n* junction phototransistors have shown a closer adherence to shot noise behavior, with proper shot dependence on the current crossing the junction. The purpose of this letter is to describe the results of this latest study.

The measurements consisted of determining the ac short-circuit noise current in a band about 400 cps wide centered at 1000 cps. The apparatus provided means for dc biasing the phototransistor under test, both in the dark and under several different conditions of illumination of constant color temperature. The measurements were made on seven *p-n* junction phototransistors carefully processed according to the best current techniques, and selected for low dc reverse saturation dark current.

The test measurements were nearly identical for all seven units, and the results can, therefore, be presented by giving detailed data for one typical unit. In Fig. 1 the ac short-circuit noise currents, for several different dc bias voltages, are plotted as functions of dc bias current. The various values of bias current were obtained by varying the illumination of the phototransistor.

The results for this particular unit show that for bias voltages under 20 v and for dc currents below about 300 μa, the short-circuit ac noise current is given within a factor of two by the theoretical shot noise current $i^2=2eI$, where e is the electronic charge and I is the dc current crossing the junction. Corresponding measurements for the other six units yield data falling within a factor of two of the plotted data over the entire range of dc biases and currents depicted.

The noise behavior of one of the units was measured at 5-v dc bias at dry ice temperature and was found to be not measurably different from the behavior at room temperature. This result was expected, in view of the known temperature independence of shot noise.

For dc currents above three or four hundred μa and for bias voltages higher than 20 v, the measured noise shows considerable excess over shot noise. In some cases this noise exhibits a non-random appearance on a CRO screen. It is conjectured that this excess noise is the persisting vestige of the $1/f$ noise observed in earlier devices, and that at higher measuring frequencies it, too, would disappear, leaving shot noise (or, what is the same thing, photon fluctuation noise in diodelike devices carrying photocurrents) as the sole mechanism necessary to account for the noise observed.

We thank Mr. H. C. Montgomery and Mr. M. A. Clark for their advice in the construction of our test equipment and discussions of the interpretation of our results.

[1] R. M. Ryder and R. J. Kircher, Bell System Tech. J. 28, 367–400 (1949).
[2] H. C. Montgomery, Bell System Tech. J. 950–975 (1952).
[3] H. C. Montgomery and M. A. Clark, J. Appl. Phys. 24, 1337 (1953).
[4] J. N. Shive, Proc. Inst. Radio Engrs. 40, 1410 (1952).

On an Indeterminate Integral in Einstein's Theory of the Viscosity of a Suspension

M. Mooney
General Laboratories, United States Rubber Company, Passaic, New Jersey
(Received October 26, 1953)

IN a recent article by Nobuhiko Saitô[1] there is a discussion of an indeterminate improper integral occurring in the Einstein[2] theory of the viscosity of a suspension of spherical particles.

The integral in question gives the mean increment to the local rate of strain at an arbitrary point in the liquid due to the presence of the suspended particles. Thus, the integral is the sum of the perturbations resulting from all the particles in the liquid distributed at random throughout space. However, since no particle of radius a can have its center located within a sphere of radius a about a point *in the liquid*, this central spherical space of radius a must be excluded from the volume of integration. If this is done, and if the outer boundary of integration is a large sphere, as is the case in Einstein's treatment, then the integral vanishes. On the other hand, if the outer boundary of integration consists of two parallel planes, the integral is nonzero and leads to results of the form

$$A=A^*/(1-\varphi), \tag{1}$$

where A^* is the mean x-strain rate of the suspension, A is the mean x-strain rate of the liquid phase of the suspension, and φ is the partial volume of the solid phase.

That this equation is correct can be shown by a very simple but rigorous argument. Let u be the x velocity at any point in the suspension, with the rate of strain within a particle indicated by $(\partial u/\partial x)_i$, and the exterior rate of strain, in the liquid, indicated by $(\partial u/\partial x)_e$. The derivative without subscript, $\partial u/\partial x$, will denote the rate of strain either in a particle or in the liquid.

Let us now take the line integral of $\partial u/\partial x$ along a straight line parallel to the x axis, over the distance x_2-x_1. Then in the limit as x_2-x_1 becomes infinitely large in comparison with the mean distance between particles, we have by definition

$$\int_{x_1}^{x_2}\frac{\partial u}{\partial x}dx=A^*(x_2-x_1). \tag{2}$$

Also, we have, identically,

$$\int_{x_1}^{x_2}\frac{\partial u}{\partial x}dx=\int_{x_1}^{x_2}\left(\frac{\partial u}{\partial x}\right)_e dx_e+\int_{x_1}^{x_2}\left(\frac{\partial u}{\partial x}\right)_i dx_i, \tag{3}$$

where the subscript e applied to dx signifies that the integral extends only over those segments of the line of integration which lie exterior to the particles, while the subscript i applied to dx has similar significance for the line segments within the particles.

The total length of line segments within the liquid is $(1-\varphi)\times(x_2-x_1)$. Hence, by the definition of A,

$$\int_{x_1}^{x_2}\left(\frac{\partial u}{\partial x}\right)_e dx = A(x_2-x_1)(1-\varphi). \tag{4}$$

Furthermore, since the particles are rigid,

$$\left(\frac{\partial u}{\partial x}\right)_i \equiv 0. \tag{5}$$

Equation (1) follows directly from the last four equations.

In considering the application of Eq. (1) in the latter part of Einstein's analysis it may be noted that here a slight error comes in. Since two particle centers cannot approach closer than the distance 2a, the line segments indicated by dx_e should properly exclude not merely the solid particles, but spheres of radius 2a about the particle centers. However, the error involved here is of the order φ^2 and is negligible in the first approximation.

Saitô is of the opinion that the difficulty of the indeterminate integral results from the neglect of the inertia terms in the Navier-Stokes equations; and he apparently hopes to publish later a more exact solution of the problem based on Oseen's equations. It is the present writer's opinion, based largely on the foregoing analysis, that the inertia terms are not important in the problem. Rather, the confusion and uncertainty seem to be connected with certain boundary properties of the solution employed for the flow near a single particle. This solution, given by Einstein and used by all subsequent writers, has the property that the net perturbation flow in any direction vanishes when integrated over a plane surface, normal to any one of the rate of strain principle axes, but does not vanish when integrated over a spherical surface, no matter how large the sphere. It seems reasonable to believe, therefore, that in order to obtain a valid and unambiguous perturbation integral from the flow equations, it is necessary to choose an outer boundary of integration such that the net perturbation flow due to a single sphere vanishes; or, if a spherical or other boundary is chosen arbitrarily, the solution used for a single particle must satisfy this same boundary condition on the chosen boundary. A more rigorous but perhaps unnecessary boundary condition on the single-particle solution would be that the perturbation resulting from the single particle must vanish at all points on the chosen boundary.

[1] N. Saitô, J. Phys. Soc. Japan **7**, 447–450 (1952).
[2] A. Einstein, Ann. phys. **19**, 289–306 (1906); **34**, 591–592 (1911).

A Remark on Stability

KENNETH S. MILLER
New York University, New York, New York
(Received September 8, 1953)

IF one has a linear system N in which the output $u(t)$ is related to the input by the linear differential operator L,

$$Lu(t)=v(t),$$

then the stability of the linear system is one of the important problems that arise in any investigation of N. There are a variety of ways[1] of defining stability and a number of forms in which the various criteria can be worded. It can be stated, for example, in terms of the analyticity of the Laplace transform of the system function of N, the character of the fundamental solutions of $Lu=0$, the properties of the impulsive response, or one-sided Green's function[2] of N. In this note we would like to prove a theorem, which might be useful in further stability studies, relating the integrability of the Green's function to that of a fundamental system of solutions of $Lu=0$.

Theorem. Let

$$L=p_0(t)\frac{d^n}{dt^n}+p_1(t)\frac{d^{n-1}}{dt^{n-1}}+\cdots+p_n(t)$$

be a linear differential operator with coefficients $p_i(t)$ of class C^{n-i}, $i=0, 1, \cdots, n$ in the semi-infinite interval $I:[0,\infty)$. Let $p_0(t)>0$ in any closed bounded subset of I. Let $H(t,y)$ be the one-sided Green's function for L. Let $\{\phi_k(t)\}$ be n linearly independent solutions of $Lu=0$. Then the statements

$$\int_0^\infty |H(t,y)|\,dt<\infty \text{ for all } y \text{ in } I, \tag{a}$$

$$\int_0^\infty |\phi_k(t)|\,dt\leqq M<\infty,\ k=1, 2, \cdots, n \tag{b}$$

are coextensive.

Before proving the theorem we might note that in some stability criteria (see reference 1) the integrability of the one-sided Green's function is relative to the second argument. One might conjecture whether the above theorem is true if we replace (a) by

$$\int_0^\infty |H(t,y)|\,dy<\infty \text{ for all } t \text{ in } I.$$

The answer is no. For let $Lu=u'-u$. Then $H(t,y)=e^{t-y}$ and $\phi(t)=e^t$. Hence, while (a) is true, $\int_0^\infty |H(t,y)|\,dy=e^t<\infty$ for all t in I; we have $\int_0^\infty |\phi(t)|\,dt=\int_0^\infty e^t dt=\infty$.

If one wishes to integrate with respect to the second argument, one could replace $H(t,y)$ by the adjoint Green's function $H^+(t,y)$. In this case, conditions (a) and (b) of the theorem would become

$$\int_0^\infty |H^+(y,t)|\,dt<\infty \text{ for all } y \text{ in } I, \tag{a}$$

$$\int_0^\infty |\phi_k(t)|\,dt\leqq M<\infty,\quad k=1, 2, \cdots, n. \tag{b}$$

We now proceed to prove the theorem. Let $\{\phi_k^+(t)\}$ be n linearly independent solutions of $L^+u=0$ so chosen that $\phi_i^+(0)=\delta_{1i}$. Then we may write

$$H(t,y)=\sum_{i=1}^n \phi_i^+(y)\phi_i(t).$$

1. (b) implies (a). This is trivial for

$$\int_0^\infty |H(t,y)|\,dt=\int_0^\infty \Big|\sum_{i=1}^n \phi_i^+(y)\phi_i(t)\Big|\,dt\leqq\sum_{i=1}^n |\phi_i^+(y)|\int_0^\infty |\phi_i(t)|\,dt$$

$$\leqq M\sum_{i=1}^n |\phi_i^+(y)|<\infty \text{ for all } y \text{ in } I.$$

2. (a) implies (b). If we let $y=0$, then

$$M_0=\int_0^\infty |H(t,0)|\,dt=\int_0^\infty |\phi_1(t)|\,dt,$$

and there exists at least one solution $\phi_1(t)$ of $Lu=0$ with the property (b).

The proof now proceeds by induction. We shall show that if there exist k linearly independent solutions of $Lu=0$, say $\phi_1, \phi_2, \cdots, \phi_k$ with property (b), then there exist $k+1$ linearly independent solutions of $Lu=0$ with this property.

Along with $\phi_1, \phi_2, \cdots, \phi_k$ choose $n-k$ solutions of $Lu=0$, say $\phi_{k+1}, \phi_{k+2}, \cdots, \phi_n$, such that $\phi_1, \phi_2, \cdots, \phi_n$ are linearly independent. Let $\phi_i^+(t)$ be the corresponding solutions of the adjoint equation. Then

$$\int_0^\infty |H(t,y)|\,dt=\int_0^\infty |\phi_1^+(y)\phi_1(t)+\cdots+\phi_n^+(y)\phi_n(t)|\,dt,$$

with $\int_0^\infty |\phi_j(t)|\,dt\leqq M$, $j=1, 2, \cdots, k$. Let

$$K(t,y)=\sum_{j=k+1}^n \phi_j^+(y)\phi_j(t),$$

and choose a y_0 such that $K(t,y_0)\not\equiv 0$. By the hypothesis (a) of the theorem

$$\int_0^\infty |H(t,y_0)|\,dt=M'<\infty,$$

and

$$\int_0^\infty |H(t,y_0)-K(t,y_0)|\,dt \leqq \sum_{j=1}^{k} |\phi_j^+(y_0)| \int_0^\infty |\phi_j(t)|\,dt = M_1 < \infty.$$

Hence,

$$M'+M_1 \geqq \int_0^\infty |K(t,y_0)|\,dt,$$

and $\phi_1(t), \phi_2(t), \cdots, \phi_k(t), K(t,y_0)$ form $k+1$ linearly independent solutions of $Lu=0$ with the property (b).

[1] L. A. Zadeh, J. Appl. Phys. **22**, 402–405 (1951).
[2] K. S. Miller, J. Appl. Phys. **22**, 1054–1057 (1951).

Titanium Observed by Electron Diffraction

SHIGETO YAMAGUCHI
Scientific Research Institute, Ltd., Hongo, Tokyo, Japan
(Received November 3, 1953)

A COMPARATIVELY high voltage (about 200 kv) was applied for driving electrons in order to obtain transmission patterns of titanium. Kroll titanium (99.5) was filed with a flat steel file. (It was impossible to section titanium by means of microtomy.) The titanium filings thus prepared were placed on a wire gauze to be observed by electron diffraction (camera length, 50 cm, wavelength of electrons, 0.028A). Figures 1 and 2 are the

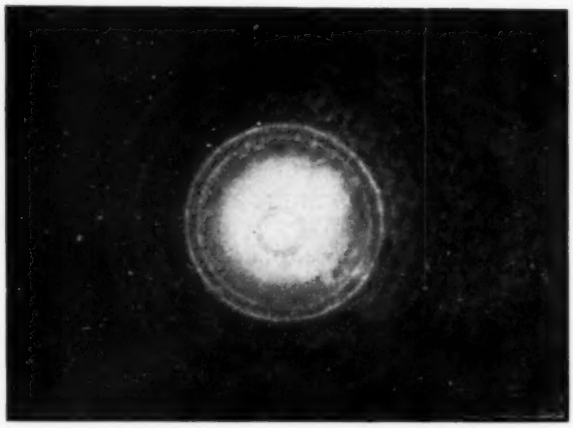


FIG. 1. Transmission pattern from a flake of hexagonal titanium. Camera length, 50 cm. Wavelength, 0.028A. Positive enlarged about two times.

diffraction patterns obtained. Table I shows the interplanar spacings and diffraction intensity measured in Fig. 1.

It is seen in Fig. 1 and Table I that the (002) diffraction of titanium is very weak in the case of electron diffraction, whereas the corresponding diffraction is comparatively strong in the case of x-ray. This fact is recognizable in the pattern of Fig. 2, in which the (100) and its equivalent diffraction spots are arranged nearly on the corners of hexagon, i.e., the incident beam is nearly normal to the (001) planes. This implies that titanium crystals are cleaved

TABLE I. d, interplanar spacings of titanium observed in the present study. d^*, interplanar spacings observed by x-ray. ( ), diffraction intensity approximated on arbitrary scale of 10.

hkl	d	d^*
100	2.54A (5)	2.54A (2.7)
002	2.34 (0.5)	2.34 (2.0)
101	2.24 (10)	2.23 (10)
102	1.73 (4)	1.72 (1.3)
110	1.474 (7)	1.470 (1.3)
103	1.323 (3)	1.330 (1.3)
200	1.275 (2)	1.275 (0.1)
112	1.243 (6)	1.248 (1.1)
201	1.227 (1)	1.230 (0.5)
202	1.126 (1)	...

[a] Hanawalt, Rinn, and Frevel, Ind. Eng. Chem. Anal. Ed. **10**, 457 (1938).

through the (001) planes of the hexagonal lattice on being filed. The filed lustrous surface of a titanium ingot gave rise to a reflexion pattern of Fig. 3, in which the (002) diffraction is also very weak.

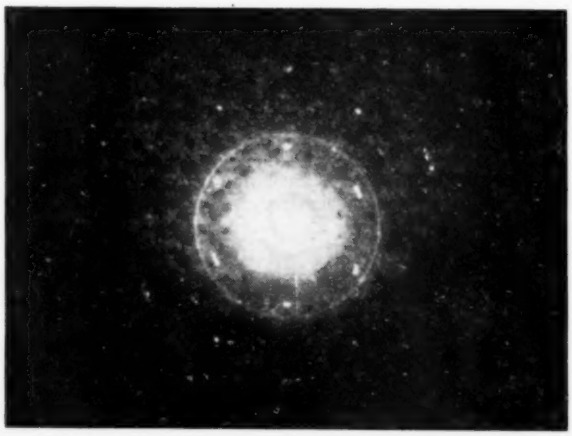


FIG. 2. Transmission pattern in which the incident beam is nearly normal to the (001) planes of titanium crystal.

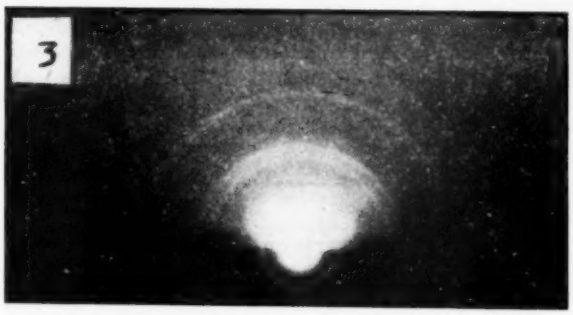


FIG. 3. Reflexion pattern from the surface of titanium ingot finished with a flat steel file.

The interplanar spacings and diffraction intensity determined here are useful as reference data for identification of thin layers of titanium, e.g., of electrodeposit, evaporated film, etc., carried out by electron diffraction.